Integrating GWASs and Human Protein Interaction Networks Identifies a Gene Subnetwork Underlying Alcohol Dependence

Shizhong Han,[1,2,*] Bao-Zhu Yang,[1] Henry R. Kranzler,[3,4] Xiaoming Liu,[5] Hongyu Zhao,[6] Lindsay A. Farrer,[7] Eric Boerwinkle,[5,8] James B. Potash,[2] and Joel Gelernter[1,9]

Despite a significant genetic contribution to alcohol dependence (AD), few AD-risk genes have been identified to date. In the current study, we aimed to integrate genome-wide association studies (GWASs) and human protein interaction networks to investigate whether a subnetwork of genes whose protein products interact with one another might collectively contribute to AD. By using two discovery GWAS data sets of the Study of Addiction: Genetics and Environment (SAGE) and the Collaborative Study on the Genetics of Alcoholism (COGA), we identified a subnetwork of 39 genes that not only was enriched for genes associated with AD, but also collectively associated with AD in both European Americans ($p < 0.0001$) and African Americans ($p = 0.0008$). We replicated the association of the gene subnetwork with AD in three independent samples, including two samples of European descent ($p = 0.001$ and $p = 0.006$) and one sample of African descent ($p = 0.0069$). To evaluate whether the significant associations are likely to be false-positive findings and to ascertain their specificity, we examined the same gene subnetwork in three other human complex disorders (bipolar disorder, major depressive disorder, and type 2 diabetes) and found no significant associations. Functional enrichment analysis revealed that the gene subnetwork was enriched for genes involved in cation transport, synaptic transmission, and transmission of nerve impulses, all of which are biologically meaningful processes that may underlie the risk for AD. In conclusion, we identified a gene subnetwork underlying AD that is biologically meaningful and highly reproducible, providing important clues for future research into AD etiology and treatment.

Introduction

Alcohol dependence (AD) (MIM 103780) is a common psychiatric disorder that is costly to individuals and to society in the United States and throughout the world. Although the etiology of AD is complex, family and twin studies provide strong evidence for a genetic contribution to its risk, with the estimated heritability ranging from 40% to 60%.[1–4] Linkage and candidate gene association studies have identified several genes that contribute to AD.[5] In recent years, genome-wide association studies (GWASs) have become a popular approach to identify common variants associated with human complex disorders including AD.[6–9] However, most published GWASs of AD relied on analysis at the SNP level; few findings have been genome-wide significant and replicated in independent samples. More-sophisticated analyses of existing GWAS data, rather than SNP-level analysis, has the potential to enhance the identification of true genetic signals, advance our understanding of its biological underpinnings, and contribute to the development of innovative diagnostic and therapeutic strategies.

Gene-set-based analysis of GWASs has been proposed to examine groups of functionally related genes, each of which has too small an effect to be detected individually, but which might be detectable when examined jointly.[10,11] Gene-set-based analysis is based on the premise that genes do not work in isolation; instead, genes that belong to the same molecular networks and cellular pathways are often involved in disease susceptibility.[12,13] Compared with SNP-level analysis, gene-set-based analysis can potentially detect joint effects or high-order interactions among genes, increase the reproducibility of significant findings, and provide more insights into disease biology. Pathway-based analysis of GWASs is one of several methods of gene-set-based analysis, which aims to detect significantly enriched gene sets from predefined canonical pathways or functional annotations. As a successful example of a pathway-based approach in GWASs, Wang et al. identified a significant pathway for Crohn disease (MIM 266600)—the *Il12/Il23* pathway—that was replicated in three independent samples.[14] This pathway included multiple genes that were not detected in any single GWAS, but were confirmed as risk genes for Crohn disease through meta-analysis of several GWASs.[15] We applied similar methods, using GWASs and pathway-based analysis, to study risk genes for cocaine dependence[16] and opioid dependence (MIM 61004).[17]

[1]Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06510, USA; [2]Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, IA 55242, USA; [3]Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA; [4]VISN 4 MIRECC, Philadelphia VAMC, Philadelphia, PA 19104, USA; [5]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [6]Departments of Biostatistics and of Genetics, Yale University School of Medicine, New Haven, CT 06510, USA; [7]Departments of Medicine (Biomedical Genetics), Neurology, Ophthalmology, Genetics & Genomics, Biostatistics, and Epidemiology, Boston University Schools of Medicine and Public Health, Boston, MA 02118, USA; [8]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [9]Departments of Genetics and of Neurobiology, Yale University School of Medicine, New Haven, CT 06510, USA
*Correspondence: shizhong-han@uiowa.edu
http://dx.doi.org/10.1016/j.ajhg.2013.10.021.

Although pathway-based analysis of GWASs has successfully identified gene sets associated with complex disorders, these analyses as traditionally performed are restricted to gene sets defined by prior information that is often incomplete. In addition, pathway-based analysis cannot detect genes that work across pathways. Network-based analysis of GWASs has been proposed as an alternate approach to pathway-based analysis to search for groups of functionally related genes within the context of gene networks. One network-based approach uses information from a human protein interaction network (HPIN) to search for a group of genes whose protein products interact with one another and which may collectively contribute to disease risk.[18] Network-based approaches are more flexible in defining gene sets, better able to detect genes that work across pathways, and less biased by prior knowledge than are pathway-based analysis.

In the current study, we sought to integrate GWASs and HPINs to identify a gene subnetwork that contributes to AD risk. Specifically, we aimed to (1) conduct a dense module search within the HPIN to identify a gene subnetwork enriched for AD-associated genes by using two GWAS data sets from the Study of Addiction: Genetics and Environment (SAGE) and the Collaborative Study on the Genetics of Alcoholism (COGA); (2) assess the association of the gene subnetwork with AD across five independent case-control samples; (3) examine the role of the same gene subnetwork in other human complex disorders; and (4) complete functional enrichment analysis to evaluate whether genes in the subnetwork share specific functional features.

Subjects and Methods

Study Descriptions

We first used two discovery GWAS data sets, SAGE and COGA, to identify a gene subnetwork enriched for AD-associated genes. To replicate significant findings, we tested the gene subnetwork in three independent case-control samples derived from the GWAS of Alcohol Use and Alcohol Use Disorder in Australian Twin-Families (OZ-ALC) and from two GWAS data sets of the Yale University School of Medicine and the University of Pennsylvania Perelman School of Medicine (Yale-Penn). To evaluate its specificity, we also examined the same gene subnetwork in three other human complex disorders: bipolar disorder (BD [MIM 125480]), major depressive disorder (MDD [MIM 608516]), and type 2 diabetes (T2D [MIM 125853]). The basic information for these GWAS data sets is summarized in Table S1 available online. In the current study, AD was defined based on DSM-IV criteria. The control subjects did not meet the DSM-IV criteria of AD, but some were dependent on other substances, as we described below for each GWAS data set. All of the work done in this paper was approved by local institutional review boards and proper informed consent was obtained. The following sections provide detailed information for each GWAS data set used in the current study.

Discovery Cohort: SAGE and COGA

SAGE aims to identify genetic risk factors and the interplay of genes and environmental factors for addiction, as described previously.[7] The current study included 2,332 unrelated European Americans (EAs) and 1,088 unrelated African Americans (AAs). In the SAGE data set, controls were defined as subjects without dependence on any substance, including cannabis, alcohol, cocaine, opioids, nicotine, and other substances. The COGA samples for GWAS were drawn from an ongoing, family-based study of AD that included subjects from seven sites around the United States.[6,19] After excluding the subjects that overlapped with those in SAGE, 478 independent subjects were retained for COGA. Because both COGA and SAGE samples were genotyped on the same ILLUMINA Human 1 M platform, we merged the 478 independent subjects from COGA with SAGE, retaining 2,670 EAs (1,453 AD and 1,217 controls) and 1,150 AAs (708 AD and 442 controls) for further analysis. In the COGA data set, controls are subjects who have consumed alcohol but do not meet any definition of AD or alcohol abuse or any DSM-III-R or DSM-IV definition of abuse or dependence on other drugs (except nicotine).

Replication Cohorts: OZ-ALC and Yale-Penn GWAS

The OZ-ALC sample was drawn from Australian twin-family samples, which have been used for linkage analysis of smoking-related traits and alcohol-use disorders.[20,21] A family-based GWAS of alcohol use and AD was constructed from these families.[9] We applied the raw data files from dbGaP that included 6,145 subjects with both phenotype and genotype information, and most subjects had relatives. To replicate the significant findings in a case-control sample set, we randomly selected 2,228 unrelated subjects (795 AD and 641 controls) from the total samples for the association analysis. In the OZ-ALC sample, controls were defined as subjects without dependence on alcohol or nicotine and without depression.

The Yale-Penn GWAS data set included individuals recruited for genetic studies of cocaine, opioid, and alcohol dependence, as described previously.[16,17,22] Subjects were interviewed with the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) to derive diagnoses for lifetime psychiatric and substance use disorders based on DSM-IV criteria.[23,24] DNA samples of 5,799 subjects interviewed via the SSADDA were genotyped on the Illumina HumanOmni1-Quad v1.0 microarray. Because the genotyped subjects include affected sibling pairs, we randomly selected 3,008 unrelated AAs (1,641 AD and 1,367 controls) and 1,407 unrelated EAs (1,086 AD and 321 controls) for the association analysis. In AAs, controls were defined as subjects without dependence on any substance, including cannabis, alcohol, cocaine, opioids, nicotine, and other substances. In EAs, controls were defined as subjects without dependence on alcohol, though they could be dependent on other substances, such as cocaine, opioids, and nicotine.

GWAS Data Sets of Other Human Complex Disorders

The GWAS data set of BD included 1,190 cases from the Bipolar Genome Study (BiGS) and 401 controls, referred to as the "TGEN" sample, which was described previously.[25] The Genetics of Recurrent Early-onset Depression (GenRED) GWAS included 1,020 MDD cases, with 1,636 controls from the Molecular Genetics of Schizophrenia (MGS) study. Details of the GenRED GWAS were described elsewhere.[26] Two GWAS data sets of T2D included 9,325 unrelated EAs (626 T2D cases and 8,699 controls) and 3,096 unrelated AAs (539 T2D cases and 2,557 controls), which were derived from the Atherosclerosis Risk in Communities (ARIC) study.[27] T2D was defined as fasting glucose $\geq$ 126 mg/dl,

nonfasting glucose ≥ 200 mg/dl, self-reported physician diagnosis of T2D, or treatment for T2D.

Data Quality Control

Each GWAS data set was cleaned before analysis via PLINK.[28] Samples and SNPs were excluded from analysis based on predetermined quality control (QC) metrics, including sample call rate ≤ 95%, SNP call rate ≤ 95%, minor allele frequency (MAF) ≤ 0.01 in controls, and p values of Hardy-Weinberg Equilibrium (HWE) tests ≤ 1×10^{-6} for controls. We used EIGENSOFT to compute principal components (PC) for samples in each GWAS data set by using pruned SNPs that were in low linkage disequilibrium (LD) with one another.[29] We removed outlier subjects from the analysis; these were defined as subjects whose ancestry was at least three standard deviations from the mean on one of the two largest PCs.

Statistical Analysis

SNP-Level Association Tests

We used PLINK for the SNP-trait association test, with sex, age, and the top three PCs as covariates under a log-additive genetic model. We evaluated the possibility of population stratification or other systemic biases by using the quantile-quantile (QQ) plots based on p values of autosomal SNPs.

Identification of Modules Enriched for Genes Associated with AD

By using an R package, dmGWAS,[30] we conducted a dense module search within the HPIN to search for modules that were enriched with small p value genes. We assigned a SNP to a gene if it was located within the gene or up to 20 kb immediately upstream or downstream, assigning a gene-level p value to each gene using the smallest p value among all SNPs mapped to that gene. A node-weighted HPIN was then constructed by superposing the gene-level p values on a HPIN from the Protein Interaction Network Analysis platform (PINA).[31] The PINA was obtained from the PINA website, and the Bioconductor package "org.Hs.eg.db" annotation resource was used to map uniprot IDs to human official gene symbols, generating ~57,800 protein interaction pairs with human gene symbols. The dmGWAS grows the module from each node by adding the neighboring nodes that can generate the maximum increment of a module score Z_m. Specifically, $Z_m = \sum Z_i/\sqrt{k}$, where k is the number of genes in the module and Z_i is transferred from the p value according to an inverse normal distribution function. Module growth is stopped if adding neighborhood nodes does not yield an increment $> Z_m \times 0.1$.

To determine how likely it is that the investigated (index) module could be identified by chance, dmGWAS creates a background distribution by scoring 100,000 randomly selected modules with the same number of genes as the index module. The significance of the index module is calculated as the proportion of those random selected modules whose Z_m are larger than or equal to that of the identified module. To adjust for module size and make the modules directly comparable to each other, dmGWAS also calculates a normalized module score Z_n, defined as $Z_n = (Z_m - mean(Z_m(\pi)))/sd(Z_m(\pi))$, from which $Z_m(\pi)$ represents the distribution of Z_m generated by scoring 100,000 randomly selected modules with the same size as the index module.

By using the two independent samples (AAs and EAs) from the merged SAGE-COGA data set, we searched for modules enriched for AD-associated genes by a dual-evaluation strategy. We first applied a dense module search in the EA sample and selected the top 5% of ranked modules in the distribution of Z_n as candidate modules for follow up in AAs. We then tested whether the candidate modules selected in EAs significantly enriched the AD-associated genes in AAs. The modules that remained significant after Bonferroni correction in AAs were designated as final candidate modules, which were then merged to construct a subnetwork for further association analysis.

Association Analysis of the Gene Subnetwork

The enrichment of small p value genes in the identified subnetwork does not necessarily mean that the subnetwork is associated with AD. We used a permutation-based association test to evaluate empirically whether the genes in the subnetwork were collectively associated with AD or the other human complex disorders investigated in the current study. We extracted the subset of data for SNPs mapped to genes belonging to the subnetwork. We then randomly reshuffled the phenotypes and genotypes while keeping the phenotype and genotype correlation structures unchanged. In each randomly permuted data set, we calculated Z_m in exactly the same way as in the real data set. This process was repeated 10,000 times and 10,000 instances of Z_m were obtained to get the empirical distribution of Z_m under the null hypothesis. The empirical p value for the association of the subnetwork with AD was estimated by counting the proportion of the Z_m that are larger than or equal to the observed ones in real data across the 10,000 permuted data sets.

Gene Ontology Enrichment Analysis

We submitted the genes in the subnetwork to The Database for Annotation, Visualization and Integrated Discovery (DAVID) for gene ontology (GO) term enrichment analysis based on GO level four annotations.[32] Fisher's exact test was implemented in DAVID to compute the enrichment p value for each GO term, followed by Bonferroni correction for multiple testing correction.

Results

SNP-Level Association Test in SAGE-COGA Data Set

We conducted association tests for each autosomal SNP with AD in the merged SAGE-COGA data sets for the EA and AA samples separately. The QQ plot of p values of the association tests indicated minimal evidence of population stratification or other systematic bias in the merged SAGE-COGA data sets (Figure S1). Overall, no single SNPs were identified that met genome-wide significance ($p = 5.0 \times 10^{-8}$), although there were some suggestive associations. For example, in EAs, the strongest association with AD was detected for rs10914375 within *SERINC2* (MIM 614549) ($p = 1.9 \times 10^{-7}$, OR = 1.38, 95% confidence interval [CI] = 1.22–1.55). In AAs, the strongest association was found for an intergenic variant rs7182484 ($p = 3.5 \times 10^{-7}$, OR = 1.62, 95% CI = 1.34–1.94). However, the strongest signal in EAs was not replicated in AAs and vice versa. The Manhattan plot of the genome-wide association test results in AAs and EAs are shown in Figure S2.

Identification of Subnetwork Enriched for AD-Associated Genes

With the merged SAGE-COGA data set, we performed dense module searching within the HPIN to identify modules enriched for AD-associated genes. A total of

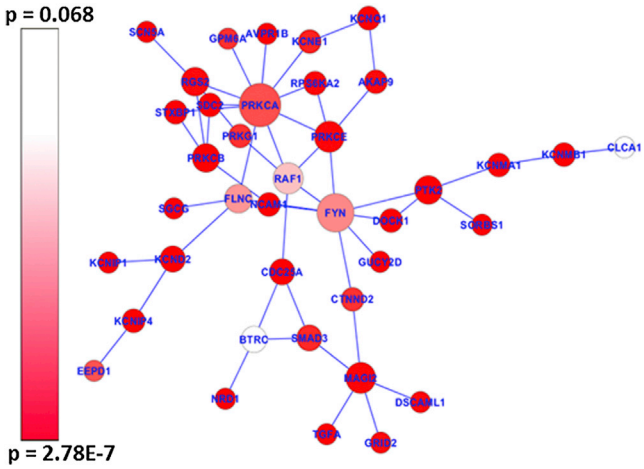

Figure 1. Gene Subnetwork Constructed with the Top 5% Modules Generated in EAs and Replicated in AAs of the Merged SAGE and COGA GWAS Data Sets

The red-white color gradient of a node is proportional to its p values. The size of a node is proportional to its degree.

8,574 modules were generated in EAs. We selected 429 top-ranked modules in the 5% upper tail of Z_n distribution and examined them in AAs for enrichment of AD-associated genes. Of the 429 modules identified in EAs, 7 were also significant in the enrichment for AD-associated genes after Bonferroni correction in AAs ($p_{corrected} < 0.05$). Table S2 shows the associated genes and statistics for each of the seven modules. The seven modules, which overlap considerably in their gene content, were then combined to construct a subnetwork including 39 nonredundant genes. Figure 1 illustrates the characteristics of the gene subnetwork. The basic information for each of the 39 genes is shown in Table S3.

Association Tests between the Gene Subnetwork and AD

We tested the cumulative evidence for association with AD for all 39 genes in the subnetwork. We observed significant associations for the gene subnetwork with AD in both EAs ($p < 0.0001$) and AAs ($p = 0.0008$) in the merged SAGE-COGA GWAS data set, suggesting that the 39-gene subnetwork was not only enriched for AD-associated genes, but also collectively associated with AD. We next asked whether the 39-gene subnetwork would show evidence of association with AD in independent samples. When we examined two European ancestry samples and one African ancestry sample, we found replication of association in all three: a European-ancestry Australian sample from OZ-ALC GWAS ($p = 0.006$) and the EA ($p = 0.001$) and AA ($p = 0.0069$) samples collected at Yale and Penn.

Association Tests between the Gene Subnetwork and Other Human Complex Disorders

We also examined whether the same gene subnetwork is associated with other human complex disorders in four independent GWAS data sets, including two psychiatric disorders (BD and MDD) in EAs and one nonpsychiatric disorder (T2D) in EAs and AAs. We found no significant evidence for association of this gene subnetwork with BD ($p = 0.53$), MDD ($p = 0.46$), or T2D (EAs, $p = 0.41$; AAs, $p = 0.25$).

GO Enrichment Analysis

This analysis revealed a number of GO terms that remained significantly enriched even after Bonferroni correction (Table 1). The enriched GO terms related to neuronal systems are particularly interesting. These include cation transport, ion transport, synaptic transmission, and transmission of nerve impulses.

Discussion

In this study, we identified a subnetwork of 39 genes that not only was enriched for genes associated with AD, but also collectively associated with AD in AAs and EAs via the merged SAGE-COGA GWAS data set. We replicated the association of the gene subnetwork with AD in each of three independent samples. We did not find any significant evidence for association of the same gene subnetwork with three other human complex disorders (BP, MDD, or T2D). We thus provided evidence that the finding is both replicable (even across populations) and, at least to some extent, specific to AD. Functional enrichment analysis revealed that the gene subnetwork was enriched for genes involved in neural processes that could underlie AD.

We sought additional evidence to evaluate whether genes in the subnetwork were previously implicated in AD. First, we examined whether these genes were reported as being associated in previous genetic association studies. We found that two genes (*NCAM1* [MIM 116930], *FYN* [MIM 137025]) were reported to be associated with AD in candidate gene association studies,[33–37] although there was also a negative report for the gene *FYN*.[37] Two other genes, *AKAP9* (MIM 604001) and *KCNMA1*(MIM 600150), were also reported in an independent GWAS on alcohol-consuming subjects.[38] In addition, *GRID2* (MIM 602368) was recently identified as a shared susceptibility gene for substance use, stress response, obesity, and hemodynamic traits.[39] Second, we examined the database Ethanol-Related Gene Resource (ERGR)[40] and found 12 genes in the subnetwork (*PRKCE* [MIM 176975], *KCNMA1*, *PTK2* [MIM 600758], *DOCK1* [MIM 601403], *SDC2* [MIM 142460], *GRID2*, *NRD1* [MIM 602651], *NCAM1*, *STXBP1* [MIM 602926], *PRKCA* [MIM 176960], *RPS6KA2* [MIM 601685], and *KCND2* [MIM 605410]) that showed changed expression levels from alcohol-related microarray gene expression studies in human, rat, or mouse models. Third, with the same ERGR database, we examined whether the subnetwork genes mapped to regions detected in human linkage studies of AD or alcohol-related behavior quantitative trait loci studies in mouse. This analysis revealed seven

Table 1. GO Enrichment Analysis of Subnetwork Genes

GO Term	Overlapping Genes[a]	p[b]	p_{adj}[c]
GO:0006812~cation transport	*KCNMA1*, *KCND2*, *CLCA1*, *FYN*, *KCNE1*, *KCNIP1*, *KCNQ1*, *SCN5A*, *KCNMB1*, *KCNIP4*, *PRKCB*	1.11×10^{-6}	5.06×10^{-4}
GO:0006811~ion transport	*KCNMA1*, *KCND2*, *CLCA1*, *FYN*, *KCNE1*, *GRID2*, *KCNIP1*, *KCNQ1*, *SCN5A*, *KCNMB1*, *KCNIP4*, *PRKCB*	2.82×10^{-6}	0.0013
GO:0007268~synaptic transmission	*PRKCA*, *KCNMA1*, *PTK2*, *KCND2*, *GRID2*, *AKAP9*, *KCNIP1*, *KCNMB1*	1.10×10^{-5}	0.005
GO:0044057~regulation of system process	*PRKCA*, *KCNMA1*, *AVPR1B*, *CTNND2*, *KCNE1*, *STXBP1*, *KCNQ1*, *SCN5A*	1.40×10^{-5}	0.0063
GO:0006796~phosphate metabolic process	*PRKCA*, *PTK2*, *RPS6KA2*, *FYN*, *TGFA*, *RAF1*, *AKAP9*, *PRKG1*, *PRKCE*, *CDC25A*, *PRKCB*, *GUCY2D*	2.70×10^{-5}	0.012
GO:0019226~transmission of nerve impulse	*PRKCA*, *KCNMA1*, *PTK2*, *KCND2*, *GRID2*, *AKAP9*, *KCNIP1*, *KCNMB1*	3.11×10^{-5}	0.014
GO:0006873~cellular ion homeostasis	*PRKCA*, *KCNMA1*, *KCND2*, *AVPR1B*, *KCNE1*, *GRID2*, *KCNQ1*, *PRKCB*	4.75×10^{-5}	0.021
GO:0008016~regulation of heart contraction	*PRKCA*, *AVPR1B*, *KCNE1*, *KCNQ1*, *SCN5A*	4.92×10^{-5}	0.022

[a]Subnetwork genes included in the GO term.
[b]Fisher exact test p value.
[c]Bonferroni correction adjusted p value.

genes (*PTK2, RGS2* [MIM 600861], *NCAM1, STXBP1, SCN5A* [MIM 600163], *SORBS1* [MIM 605264], and *CDC25A* [MIM 116947]) with linkage evidence for AD or alcohol-related behavior.

It is notable that 7 genes of the 39 identified encode potassium channels or their interacting proteins. Potassium channel genes are particularly attractive, because potassium channels are regulated by alcohol and are key elements of behavioral tolerance to alcohol in both invertebrates and mammals.[41] Moreover, in an independent GWAS, two potassium channel genes, *KCNMA1* and *KCNQ5* (MIM 607357), were reported to be associated with AD, which further supports the potential role of potassium channel genes in AD.[38] In our GWAS of opioid dependence, potassium channel genes were also very prominent among the key findings.[17] Because potassium channels have been major therapeutic targets for drug discovery,[42] further research on these channels may contribute to the development of new drugs for AD treatment.

We performed additional analyses of the gene subnetwork with AD. First, because the gene subnetwork was selected on the basis of protein interactions that might include nonlinear interactions, we tested the gene subnetwork via the SNP-set (Sequence) Kernel Association Test (SKAT), which employs a statistical framework that accounts for both linear and nonlinear interactions.[43] We observed significant associations of the gene subnetwork with AD in the SAGE-COGA EA ($p = 0.00069$), OZ-ALC ($p = 0.048$), and Yale-Penn AA ($p = 0.007$) samples, reinforcing the genetic associations detected by the smallest p value method. We did not identify significant associations in the SAGE-COGA AA ($p = 0.12$) and Yale-Penn EA ($p = 0.18$) samples. We suspect that the nonsignificant results in the SAGE-COGA AA and Yale-Penn EA samples were attributable to the lower statistical power of SKAT in these two samples given their smaller sample sizes and the greater statistical noise that may exist because SKAT tested all SNPs within the gene subnetwork. Second, given the heterogeneous nature of the DSM-IV AD diagnosis, we attempted to investigate whether the gene subnetwork was associated more specifically with some related clinical subphenotypes, such as the presence of alcohol tolerance or withdrawal symptoms. Our analysis did not reveal any stronger associations for these two subgroups than for the AD diagnosis (Table S4). Third, analysis of data stratified by sex did not yield greater statistical associations for either males or females, suggesting no sex-specific effects for the gene subnetwork (Table S5).

Our results highlight several advantages of network-based analysis of GWASs over conventional analytical strategies to uncover susceptibility genes for AD. First, network-based analysis may identify genes that could be missed by traditional univariate analytical approaches. None of the 39 genes in the subnetwork reached genome-wide significance ($p = 5 \times 10^{-8}$) individually and therefore were unlikely to attract attention in SNP-based GWASs. Table S6 shows the univariate SNP analysis results with nominal significance for the 39 gene set in the SAGE-COGA EA sample. Only six genes (*NRD1*, *CDC25A*, *KCNIP1*, *NCAM1*, *SGCG* [MIM 608896], and *GUCY2D* [MIM 600179]) contained SNPs with a significance level $\leq 1 \times 10^{-4}$ in the discovery EAs from the SAGE-COGA data set. We identified these 39 genes in the context of a HPIN, suggesting that although each could be modestly associated with AD, they may collectively be detectable in AD susceptibility. Second, network-based analysis may provide more insight into disease biology. Functional enrichment analysis of the genes in the subnetwork revealed a number of functional annotations that are consistent with our prior understanding of the neurobiology of AD. Thus, identification of possible risk genes on the basis of their inclusion in the subnetwork may provide important clues to future research into the etiology of AD and potential targets for its treatment. Third, the significant association between the subnetwork and AD was consistently replicated in three independent data sets, providing

convergent validity for the findings and suggesting potentially higher reproducibility for this kind of system-level analysis. Fourth, there was no evidence of association between the subnetwork and other complex human diseases, providing divergent validation for the findings.

Our current study also has some limitations. For example, in the subnetwork construction stage, we arbitrarily selected modules that were ranked in the top 5% for EAs and further remained significant after Bonferroni correction in AAs via the SAGE-COGA data set. Hence, it is likely that some modules that contribute to AD susceptibility but did not meet our module selection criteria could have been missed. In addition, we used the smallest p value method to assign the gene-level p values. Under these circumstances, it is likely that the association signals for genes with multiple independent risk variants were not well captured.

In conclusion, through the integrated analysis of GWAS and HPIN, we identified a subnetwork implicated in AD that is biologically meaningful and highly reproducible. The genes included in the subnetwork may provide future targets for research into the etiology and treatment of AD.

Supplemental Data

Supplemental Data include Supplemental Acknowledgments, two figures, and six tables and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

The authors are grateful to the volunteer families and individuals who participated in this research study. The authors would like to thank Peilin Jia for helping in the use of dmGWAS R package. The authors are indebted to the two anonymous reviewers for their constructive comments. This study was supported by the Alcoholic Beverage Medical Research Foundation (ABMRF) grant (S.H.). This study was also supported by National Institutes of Health grants R01 DA12690, R01 DA12849, RC2 DA028909, R01 DA18432, R01 AA11330, R01 AA017535, and K01 DA24758, and by the VA Connecticut REAP center, a VA MERIT grant, VA National Center for PTSD Research, and the VA Connecticut and VISN4 MIRECC Centers. Genotyping services for a part of the Yale-Penn GWAS were provided by the Center for Inherited Disease Research (CIDR) and Yale University (Center for Genome Analysis). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University (contract number N01-HG-65403). The SAGE, COGA, and OZ-ALC GWAS data sets in this study were obtained from dbGaP. Detailed funding and support information for these GWAS data sets are in Supplemental Acknowledgements. H.R.K. has been a paid consultant or advisory board member for Alkermes, Lilly, Lundbeck, Pfizer, and Roche and is a member of the Alcohol Clinical Trials Initiative (ACTIVE) of the American Society of Clinical Psychopharmacology, which is supported by AbbVie, Lilly, Lundbeck, and Pfizer.

Received: August 14, 2013
Revised: October 14, 2013
Accepted: October 21, 2013
Published: November 21, 2013

Web Resources

The URLs for data presented herein are as follows:

Bioconductor, http://www.bioconductor.org
dbGaP, http://www.ncbi.nlm.nih.gov/gap
dmGWAS, http://bioinfo.mc.vanderbilt.edu/dmGWAS.html
EIGENSTART, http://www.hsph.harvard.edu/alkes-price/software/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/
SKAT and SKAT-O, http://www.hsph.harvard.edu/xlin/software.html

References

1. Kendler, K.S., Heath, A.C., Neale, M.C., Kessler, R.C., and Eaves, L.J. (1992). A population-based twin study of alcoholism in women. JAMA *268*, 1877–1882.
2. Kendler, K.S., Prescott, C.A., Neale, M.C., and Pedersen, N.L. (1997). Temperance board registration for alcohol abuse in a national sample of Swedish male twins, born 1902 to 1949. Arch. Gen. Psychiatry *54*, 178–184.
3. Heath, A.C., Bucholz, K.K., Madden, P.A., Dinwiddie, S.H., Slutske, W.S., Bierut, L.J., Statham, D.J., Dunne, M.P., Whitfield, J.B., and Martin, N.G. (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.
4. Prescott, C.A., and Kendler, K.S. (1999). Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. Am. J. Psychiatry *156*, 34–40.
5. Gelernter, J., and Kranzler, H.R. (2009). Genetics of alcohol dependence. Hum. Genet. *126*, 91–99.
6. Edenberg, H.J., Koller, D.L., Xuei, X., Wetherill, L., McClintick, J.N., Almasy, L., Bierut, L.J., Bucholz, K.K., Goate, A., Aliev, F., et al. (2010). Genome-wide association study of alcohol dependence implicates a region on chromosome 11. Alcohol. Clin. Exp. Res. *34*, 840–852.
7. Bierut, L.J., Agrawal, A., Bucholz, K.K., Doheny, K.F., Laurie, C., Pugh, E., Fisher, S., Fox, L., Howells, W., Bertelsen, S., et al.; Gene, Environment Association Studies Consortium (2010). A genome-wide association study of alcohol dependence. Proc. Natl. Acad. Sci. USA *107*, 5082–5087.
8. Frank, J., Cichon, S., Treutlein, J., Ridinger, M., Mattheisen, M., Hoffmann, P., Herms, S., Wodarz, N., Soyka, M., Zill, P., et al. (2012). Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. Addict. Biol. *17*, 171–180.
9. Heath, A.C., Whitfield, J.B., Martin, N.G., Pergadia, M.L., Goate, A.M., Lind, P.A., McEvoy, B.P., Schrage, A.J., Grant, J.D., Chou, Y.L., et al. (2011). A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. Biol. Psychiatry *70*, 513–518.
10. Wang, K., Li, M., and Bucan, M. (2007). Pathway-based approaches for analysis of genomewide association studies. Am. J. Hum. Genet. *81*, 1278–1283.
11. Wang, K., Li, M., and Hakonarson, H. (2010). Analysing biological pathways in genome-wide association studies. Nat. Rev. Genet. *11*, 843–854.

12. Xu, J., and Li, Y. (2006). Discovering disease-genes by topological features in human protein-protein interaction network. Bioinformatics *22*, 2800–2805.
13. Gandhi, T.K., Zhong, J., Mathivanan, S., Karthick, L., Chandrika, K.N., Mohan, S.S., Sharma, S., Pinkert, S., Nagaraju, S., Periaswamy, B., et al. (2006). Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat. Genet. *38*, 285–293.
14. Wang, K., Zhang, H., Kugathasan, S., Annese, V., Bradfield, J.P., Russell, R.K., Sleiman, P.M., Imielinski, M., Glessner, J., Hou, C., et al. (2009). Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn disease. Am. J. Hum. Genet. *84*, 399–405.
15. Barrett, J.C., Hansoul, S., Nicolae, D.L., Cho, J.H., Duerr, R.H., Rioux, J.D., Brant, S.R., Silverberg, M.S., Taylor, K.D., Barmada, M.M., et al.; NIDDK IBD Genetics Consortium; Belgian-French IBD Consortium; Wellcome Trust Case Control Consortium (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat. Genet. *40*, 955–962.
16. Gelernter, J., Sherva, R., Koesterer, R., Almasy, L., Zhao, H., Kranzler, H.R., and Farrer, L.A. (2013). Genomewide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. Mol. Psychiatry. Published online August 20, 2013. http://dx.doi.org/10.1038/mp.2013.99.
17. Gelernter, J., Kranzler, H.R., Sherva, R., Koesterer, R., Sun, J., Bi, J., Almasy, L., Zhao, H., and Farrer, L.A. (2013). Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. Biol. Psychiatry. Published online October 19, 2013. http://dx.doi.org/10.1016/j.biopsych.2013.08.034.
18. Baranzini, S.E., Galwey, N.W., Wang, J., Khankhanian, P., Lindberg, R., Pelletier, D., Wu, W., Uitdehaag, B.M., Kappos, L., Polman, C.H., et al.; GeneMSA Consortium (2009). Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Hum. Mol. Genet. *18*, 2078–2090.
19. Reich, T., Edenberg, H.J., Goate, A., Williams, J.T., Rice, J.P., Van Eerdewegh, P., Foroud, T., Hesselbrock, V., Schuckit, M.A., Bucholz, K., et al. (1998). Genome-wide search for genes affecting the risk for alcohol dependence. Am. J. Med. Genet. *81*, 207–215.
20. Saccone, S.F., Pergadia, M.L., Loukola, A., Broms, U., Montgomery, G.W., Wang, J.C., Agrawal, A., Dick, D.M., Heath, A.C., Todorov, A.A., et al. (2007). Genetic linkage to chromosome 22q12 for a heavy-smoking quantitative trait in two independent samples. Am. J. Hum. Genet. *80*, 856–866.
21. Hansell, N.K., Agrawal, A., Whitfield, J.B., Morley, K.I., Gordon, S.D., Lind, P.A., Pergadia, M.L., Montgomery, G.W., Madden, P.A., Todd, R.D., et al. (2010). Linkage analysis of alcohol dependence symptoms in the community. Alcohol. Clin. Exp. Res. *34*, 158–163.
22. Xie, P., Kranzler, H.R., Yang, C., Zhao, H., Farrer, L.A., and Gelernter, J. (2013). Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. Biol. Psychiatry *74*, 656–663.
23. Pierucci-Lagha, A., Gelernter, J., Chan, G., Arias, A., Cubells, J.F., Farrer, L., and Kranzler, H.R. (2007). Reliability of DSM-IV diagnostic criteria using the semi-structured assessment for drug dependence and alcoholism (SSADDA). Drug Alcohol Depend. *91*, 85–90.
24. Pierucci-Lagha, A., Gelernter, J., Feinn, R., Cubells, J.F., Pearson, D., Pollastri, A., Farrer, L., and Kranzler, H.R. (2005). Diagnostic reliability of the Semi-structured Assessment for Drug Dependence and Alcoholism (SSADDA). Drug Alcohol Depend. *80*, 303–312.
25. Smith, E.N., Koller, D.L., Panganiban, C., Szelinger, S., Zhang, P., Badner, J.A., Barrett, T.B., Berrettini, W.H., Bloss, C.S., Byerley, W., et al. (2011). Genome-wide association of bipolar disorder suggests an enrichment of replicable associations in regions near genes. PLoS Genet. *7*, e1002134. http://dx.doi.org/10.1371/journal.pgen.1002134.
26. Shi, J., Potash, J.B., Knowles, J.A., Weissman, M.M., Coryell, W., Scheftner, W.A., Lawson, W.B., DePaulo, J.R., Jr., Gejman, P.V., Sanders, A.R., et al. (2011). Genome-wide association study of recurrent early-onset major depressive disorder. Mol. Psychiatry *16*, 193–201.
27. Williams, O.D. (1989). The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am. J. Epidemiol. *129*, 687–702.
28. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
29. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
30. Jia, P., Zheng, S., Long, J., Zheng, W., and Zhao, Z. (2011). dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. Bioinformatics *27*, 95–102.
31. Wu, J., Vallenius, T., Ovaska, K., Westermarck, J., Mäkelä, T.P., and Hautaniemi, S. (2009). Integrated network analysis platform for protein-protein interactions. Nat. Methods *6*, 75–77.
32. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. *4*, 3.
33. Yang, B.Z., Kranzler, H.R., Zhao, H., Gruen, J.R., Luo, X., and Gelernter, J. (2008). Haplotypic variants in DRD2, ANKK1, TTC12, and NCAM1 are associated with comorbid alcohol and drug dependence. Alcohol. Clin. Exp. Res. *32*, 2117–2127.
34. Yang, B.Z., Kranzler, H.R., Zhao, H., Gruen, J.R., Luo, X., and Gelernter, J. (2007). Association of haplotypic variants in DRD2, ANKK1, TTC12 and NCAM1 to alcohol dependence in independent case control and family samples. Hum. Mol. Genet. *16*, 2844–2853.
35. Pastor, I.J., Laso, F.J., Inés, S., Marcos, M., and González-Sarmiento, R. (2009). Genetic association between -93A/G polymorphism in the Fyn kinase gene and alcohol dependence in Spanish men. Eur. Psychiatry *24*, 191–194.
36. Schumann, G., Rujescu, D., Kissling, C., Soyka, M., Dahmen, N., Preuss, U.W., Wieman, S., Depner, M., Wellek, S., Lascorz, J., et al. (2003). Analysis of genetic variations of protein tyrosine kinase fyn and their association with alcohol dependence in two independent cohorts. Biol. Psychiatry *54*, 1422–1426.
37. Ishiguro, H., Saito, T., Shibuya, H., Toru, M., and Arinami, T. (2000). Mutation and association analysis of the Fyn kinase gene with alcoholism and schizophrenia. Am. J. Med. Genet. *96*, 716–720.
38. Kendler, K.S., Kalsi, G., Holmans, P.A., Sanders, A.R., Aggen, S.H., Dick, D.M., Aliev, F., Shi, J., Levinson, D.F., and Gejman, P.V. (2011). Genomewide association analysis of symptoms of

alcohol dependence in the molecular genetics of schizophrenia (MGS2) control sample. Alcohol. Clin. Exp. Res. *35*, 963–975.

39. Nikpay, M., Šeda, O., Tremblay, J., Petrovich, M., Gaudet, D., Kotchen, T.A., Cowley, A.W., Jr., and Hamet, P. (2012). Genetic mapping of habitual substance use, obesity-related traits, responses to mental and physical stress, and heart rate and blood pressure measurements reveals shared genes that are overrepresented in the neural synapse. Hypertens. Res. *35*, 585–591.
40. Guo, A.Y., Webb, B.T., Miles, M.F., Zimmerman, M.P., Kendler, K.S., and Zhao, Z. (2009). ERGR: An ethanol-related gene resource. Nucleic Acids Res. *37* (Database issue), D840–D845.
41. Treistman, S.N., and Martin, G.E. (2009). BK Channels: mediators and models for alcohol tolerance. Trends Neurosci. *32*, 629–637.
42. Wulff, H., Castle, N.A., and Pardo, L.A. (2009). Voltage-gated potassium channels as therapeutic targets. Nat. Rev. Drug Discov. *8*, 982–1001.
43. Wu, M.C., Kraft, P., Epstein, M.P., Taylor, D.M., Chanock, S.J., Hunter, D.J., and Lin, X. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. Am. J. Hum. Genet. *86*, 929–942.