Archival Report

Biological Psychiatry

Genome-Wide Association Study of Nicotine Dependence in American Populations: Identification of Novel Risk Loci in Both African-Americans and European-Americans

Joel Gelernter, Henry R. Kranzler, Richard Sherva, Laura Almasy, Aryeh I. Herman, Ryan Koesterer, Hongyu Zhao, and Lindsay A. Farrer

ABSTRACT

BACKGROUND: We report a genome-wide association study (GWAS) of nicotine dependence defined on the basis of scores on the Fagerström Test for Nicotine Dependence in European-American (EA) and African-American (AA) populations.

METHODS: Our sample, from the one used in our previous GWAS, included only subjects who had smoked >100 cigarettes lifetime (2114 EA and 2602 AA subjects) and an additional 927 AA and 2003 EA subjects from the Study of Addiction: Genetics and Environment project [via the database of Genotypes and Phenotypes (dbGAP)]. GWAS analysis considered Fagerström Test for Nicotine Dependence score as an ordinal trait, separately in each population and sample and by combining the results in meta-analysis. We also conducted analyses that were adjusted for other substance use disorder criteria in a single nucleotide polymorphism (SNP) subset.

RESULTS: In EAs, one chromosome 7 intergenic region was genome-wide significant (GWS): rs13225753, $p = 3.48 \times 10^{-8}$ (adjusted). In AAs, GWS associations were observed at numerous SNPs mapped to a region on chromosome 14 of >305,000 base pairs (minimal $p = 4.74 \times 10^{-10}$). Two chromosome 8 regions were associated: $p = 4.45 \times 10^{-8}$ at *DLC1* SNP rs289519 (unadjusted) and $p = 1.10 \times 10^{-9}$ at rs6996964 (adjusted for other substances), located between *CSGALNACT1* and *INTS10*. No GWS associations were observed at the chromosome 15 nicotinic receptor gene cluster (*CHRNA5-CHRNA3-CHRNB4*) previously associated with nicotine dependence and smoking quantity traits. *TSNAX-DISC1* SNP rs821722 ($p = 1.46 \times 10^{-7}$) was the most significant result with substantial contributions from both populations; we previously identified *DISC1* associations with opioid dependence. Pathway analysis identified association with nitric oxide synthase and adenosine monophosphate-activated protein kinase pathways in EAs.

CONCLUSIONS: The key risk loci identified, which require replication, offer novel insights into nicotine dependence biology.

Keywords: AMPK pathway, DISC1, DLC1, eNOS pathway, FTND, GWAS, Nicotine dependence, Population differences

http://dx.doi.org/10.1016/j.biopsych.2014.08.025

Genome-wide association study (GWAS), an important step in the identification of risk genes for complex traits, has only recently been applied to gene mapping for substance dependence (SD) traits. We previously reported risk genes identified by GWAS for cocaine, alcohol, and opioid dependence (1–4). By far, the most studied SD trait from a genetic perspective is nicotine dependence (ND), which is moderately heritable (h^2 = .48–.72 based on twin studies) (5,6). The heritability of scores on the Fagerström Test for Nicotine Dependence (FTND), a quantitative measure frequently used to measure ND (7), was estimated to be .40 to .75 (8–10). Many GWAS studies and several meta-analyses of ND-related traits have been published. The most consistent signals identified via GWAS emerge from a set of closely mapped nicotinic receptor genes on chromosome 15 (11–13). In a meta-analysis of smoking behavior GWAS in African-Americans (AAs), the only genome-wide significant (GWS) association mapped to the same cluster (14).

We used GWAS to identify genetic variants that influence risk of ND as measured by the FTND. We included European-American (EA) and AA subjects who reported having smoked at least 100 cigarettes lifetime, derived from our substance dependence GWAS sample of 4716 subjects (1–3) (Yale-Penn sample), combined with a sample of 2930 subjects from the Study of Addiction: Genetics and Environment (SAGE), available to researchers through dbGAP (Database of Genotypes and Phenotypes) application.

SEE COMMENTARY ON PAGE 423

ISSN: 0006-3223

METHODS AND MATERIALS

Subjects and Diagnostic Procedures

Our GWAS discovery sample included 2114 EA and 2602 AA subjects (after exclusion of those not meeting the exposure criterion: 308 AAs and 98 EAs reported never having smoked ≥100 cigarettes). All subjects were recruited for studies of the genetics of drug (opioid or cocaine) or alcohol dependence (1–3). The sample consisted of small nuclear families originally collected for linkage studies (primarily full sibs, half sibs, and parents, generally no more than one parent per family) and unrelated individuals. Subjects (Table S1 in Supplement 1) gave written informed consent as approved by the institutional review board at each site, and certificates of confidentiality were obtained from the National Institute on Drug Abuse and National Institute on Alcohol Abuse and Alcoholism. Subjects were administered the Semi-Structured Assessment for Drug Dependence and Alcoholism (15), in which the FTND is embedded. The FTND domains assessed by this instrument are how soon the subject smokes his first cigarette after awakening; whether the subject finds it difficult to refrain from smoking in places where it is forbidden; which cigarette the subject would least like to give up (e.g., the first cigarette in the morning); how many cigarettes the subject smokes per day; and whether the subject smokes even if ill enough to be confined to his bed [paraphrased from reference (7)].

Discovery phase analyses also included publicly available (via application) GWAS data from SAGE (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1), containing 927 AA and 2003 EA unrelated exposed individuals (Table S1 in Supplement 1). SAGE includes individuals from the Collaborative Study on the Genetics of Alcoholism (COGA) (16), the Family Study of Cocaine Dependence (FSCD) (17), and the Collaborative Genetic Study of Nicotine Dependence (COGEND) (18). The COGA sample is a set of unrelated individuals recruited in Indiana, New York, St. Louis, Connecticut, Iowa, and San Diego selected for genotyping from a larger set of 8000 subjects. COGA cases met criteria for DSM-IV alcohol dependence. FSCD contained subjects from the greater St. Louis metropolitan area; most cases met criteria for DSM-IV alcohol dependence and cocaine dependence. Control subjects were from the same communities and had consumed alcohol but had no lifetime history of dependence on any substance. A subgroup of FSCD subjects was not alcohol dependent but had a lifetime DSM-IV diagnosis of dependence on cannabis or another illicit drug. COGEND subjects were recruited in Missouri and Michigan. COGEND cases met criteria for DSM-IV alcohol and/or nicotine dependence. Control subjects were selected from the nondependent population and did not meet criteria for alcohol, nicotine, or illicit drug dependence.

Genotyping and Quality Control

Yale-Penn GWAS samples were genotyped on the Illumina HumanOmni1-Quad v1.0 microarray, including 988,306 autosomal single nucleotide polymorphisms (SNPs; Illumina, San Diego, California), at the Center for Inherited Disease Research and the Yale Center for Genome Analysis. Genotypes were called using GenomeStudio software V2011.1 and genotyping module V1.8.4 (Illumina, San Diego, California). SAGE samples were genotyped on the Illumina Human 1M array containing 1,069,796 total SNPs (Illumina). In the Yale-Penn GWAS dataset, 44,644 SNPs on the microarray and 135 individuals with call rates <98% were excluded; 62,076 additional SNPs were removed due to minor allele frequencies (MAF) <1%. After data cleaning and quality control, 5697 individuals and 889,659 SNPs remained for imputation. Additional quality control information has been reported previously (1). After applying the same quality control procedures to the SAGE sample, 39 subjects with call rates <98% were excluded and 726,191 SNPs remained for analysis.

To verify and correct the misclassification of self-reported race, we compared the GWAS data from all subjects with genotypes from the HapMap 3 (http://hapmap.ncbi.nlm.nih.gov/) reference CEU (CEPH collection), YRI (Yoruba in Ibadan, Nigeria), and CHB (Han Chinese in Beijing, China) populations. Principal components (PC) analysis was conducted in the entire GWAS sample using Eigensoft (19,20) and 145,472 SNPs that were common to the GWAS dataset and HapMap panel (after pruning the GWAS SNPs for linkage disequilibrium [R^2] >80%) to characterize the underlying genetic architecture of the samples. The first 10 PC scores were used in a k-means cluster analysis to distinguish AAs and EAs; these groups were subsequently analyzed separately. We then conducted PC analyses within the two groups and the first three PCs were used in all subsequent analyses to correct for residual population stratification.

Genotype Imputation

SNP genotype imputation was performed in the Yale-Penn and the SAGE GWAS datasets with IMPUTE2 (21) using genotyped SNPs with a minor allele frequency of >1% and the June 2011 1000 Genomes reference panel (22), which contains phased haplotypes for 1094 individuals of various ancestries: 379 of European descent (CEU, FIN (Finnish in Finland), GBR (British from England and Scotland), IBS (Iberian populations in Spain), and TSI (Toscani in Italia)), 286 of Asian descent CHB, JPT (Japanese in Tokyo, Japan), and CHS (Han Chinese South, China), 181 admixed American samples (PUR (Puerto Rican in Puerto Rico), CLM (Colombian in Medellin, Colombia), and MXL (Mexican ancestry in Los Angeles, California)), and 246 samples of African descent (ASW (African ancestry in southwest USA), LWK (Luhya in Webuye, Kenya), YRI) (22). All samples were imputed using every available sample in the reference panel, then split into AA and EA datasets based on the clustering techniques described above. We retained 18,564,419 SNPs with derived information content >.8 in at least one of the population groups. After excluding SNPs with MAF < 3% in both AAs and EAs, 11,995,908 SNPs common to both discovery datasets (11,106,284 in AAs, 7,535,791 in EAs) were included in association analyses.

Statistical Analysis Methods

Association tests were performed for SNPs with MAF >3% using linear association models embedded in generalized estimating equations to correct for correlations among related individuals (23). We modeled the FTND score as a continuous variable that was analyzed in a standard linear regression and adjusted for age, sex, and three PCs of ancestry. Although the