

# Genome-Wide Association for Nicotine Dependence and Smoking Cessation Success in NIH Research Volunteers

Tomas Drgon,<sup>1</sup> Ivan Montoya,<sup>1,2</sup> Catherine Johnson,<sup>1</sup> Qing-Rong Liu,<sup>1</sup> Donna Walther,<sup>1</sup> Dean Hamer,<sup>3</sup> and George R Uhl<sup>1</sup>

<sup>1</sup>Molecular Neurobiology Branch, National Institutes of Health–Intramural Research Program (National Institute on Drug Abuse), Baltimore, Maryland, United States of America; <sup>2</sup>Division of Pharmacotherapies and Medical Consequences of Drug Abuse, National Institute on Drug Abuse, National Institutes of Health, Rockville Maryland, United States of America; <sup>3</sup>Section on Gene Structure and Regulation, Laboratory of Biochemistry, National Institutes of Health–Intramural Research Program (National Cancer Institute), Bethesda, Maryland, United States of America

Phenotypes related to both nicotine dependence and ability to successfully quit smoking display substantial heritabilities in classical and molecular genetic studies. Twin studies suggest that some genetic components for dependence overlap with genetic components of ability to quit, but that many components do not overlap. Initial genome-wide association (GWA) studies have demonstrated haplotypes that distinguish nicotine-dependent from nondependent smokers. These haplotypes overlap partially with those that distinguish individuals who successfully quit smoking from those who were not able to quit smoking in clinical trials for smoking cessation. We now report novel genome-wide association results from National Institutes of Health research volunteers who reported smoking histories, symptoms of nicotine dependence, and ability to successfully quit smoking outside the context of a clinical trial. These results buttress data from several prior GWA studies. The data from these volunteers support the idea that previously reported studies of genes associated with smoking cessation success in clinical trial participants may also apply to smokers who are more or less able to initiate and sustain abstinence outside of clinical trial settings.

© 2009 The Feinstein Institute for Medical Research, [www.feinsteininstitute.org](http://www.feinsteininstitute.org)

Online address: <http://www.molmed.org>

doi: 10.2119/molmed.2008.00096

## INTRODUCTION

Vulnerability to dependence on an addictive substance is a complex trait with strong genetic influences that are well documented by data from family, adoption, and twin studies (1–4). Twin studies support the view that much of the heritable influence on vulnerability to dependence on addictive substances from different pharmacological classes (for example, nicotine and stimulants) is shared (2,3,5). These findings have increased interest in genes that might predispose individuals to dependence on multiple substances, such as nicotine and other drugs. For smoking, substantial data also document heritabilities of nico-

tine dependence as defined by criteria from either the DSM (*Diagnostic and Statistical Manual*) or Fagerstrom test for nicotine dependence (FTND) (6–9).

Twin-study data also document substantial heritability of the ability to successfully abstain from smoking (10,11). These data suggest that some, but not most, of these genetic influences are shared with the genetics of developing dependence on nicotine.

Genome-wide association (GWA) approaches of increasing sophistication have been developed and used to identify the specific genes and genomic variants that predispose to phenotypes related to smoking and to smoking cessation, as

well as to dependence on other addictive substances (12–17). Comparisons of data from smokers whose FTND scores indicate dependence on cigarettes with data from smokers whose FTND scores display little evidence of dependence have revealed repeated associations of moderate size within a chromosome-15 cluster of nicotinic-receptor genes (15–18). Partial data from one of these studies, which also reported modest associations at a number of additional loci, are available for comparative analyses (15).

We recently reported GWA data for success in quitting smoking in each of three independent samples of carefully monitored individuals who tried to quit smoking in the context of clinical trials that used placebo, nicotine replacement, or bupropion treatments (14). These trials, however, involved highly selected smokers who were required to fulfill many entrance criteria. Individuals whose data were analyzed were required

---

**Address correspondence and reprint requests to** George R Uhl, Molecular Neurobiology, Box 5180, Baltimore, MD 21224. Phone: 443-740-2799; Fax: 443-740-2122; E-mail: [guhl@intra.nida.nih.gov](mailto:guhl@intra.nida.nih.gov).

Submitted September 3, 2008; Accepted for publication October 1, 2008; Epub ([www.molmed.org](http://www.molmed.org)) ahead of print October 2, 2008.

to participate in detailed and lengthy follow-up. Such strict selection criteria might have rendered these trial participants unrepresentative of smokers in the general population who seek to quit smoking. Replication of many of these results in individuals who quit in community settings would add to our confidence in the generalizability of these findings outside of the somewhat artificial setting of a clinical trial.

Hamer and colleagues have assembled a research volunteer population at and around the NIH clinical center in Bethesda, Maryland (19). These individuals provided data on their histories of smoking and success at quitting smoking. We now report 500k single-nucleotide polymorphism (SNP) GWA studies of the participants of the Hamer *et al.* research volunteer population. We have targeted these analyses to seek overlaps between nominally positive results from this sample and positive results from other GWA study populations, including a 38,000 SNP GWA dataset obtained for heavy smokers who displayed FTND nicotine dependence versus smokers who did not display FTND dependence (Beirut *et al.* dataset) (15), 500–600,000 SNP GWA datasets obtained for smokers who succeeded versus those who did not succeed in at least 2 of 3 clinical trials of smoking cessation (Uhl *et al.* dataset) (14), and 600,000–1 million SNP GWA datasets obtained for individuals who were dependent on at least one illegal addictive substance compared with control individuals with little or no peak lifetime use of any addictive substance, including tobacco (Liu/Drzon *et al.* dataset) (20). We discuss the significance of the substantial overlaps in the data that we report, as well as the limitations of the samples and datasets. These data buttress previous GWA results for FTND nicotine dependence. They also document, for the first time, overlaps between the molecular genetics of smoking cessation in clinical trials versus smoking cessation in community-based research volunteer samples who quit outside of the context of a clinical trial.

## MATERIALS AND METHODS

### Samples

DNA samples were obtained from individuals who were recruited at the National Institutes of Health (NIH) clinical center for protocols based on cancer risk behaviors and personality as previously described (19). DNA samples from a total of 480 individuals from the largest self-reported ethnicity in this sample, European-Americans, were divided into (a) 13 pools ( $n = 20$  in each) of DNA from individuals who reported never smoking extensively or becoming nicotine dependent, (b) 6 pools of DNA from 120 individuals who reported lifetime nicotine dependence and current smoking, and (c) 5 pools of DNA from 100 individuals who reported having been nicotine dependent at some time in their lives but who achieved abstinence. As previously reported, individuals in these three groups described their smoking history as (a) smoking fewer than 101 cigarettes in their lives, (b) starting smoking at age  $17 \pm 4$  years, smoking for  $18 \pm 13$  years, consuming  $20 \pm 13$  cigarettes/d, and continuing to smoke when interviewed, or (c) starting smoking at age  $17 \pm 3$  years, smoking an average of  $20 \pm 13$  cigarettes/d for  $13 \pm 11$  years, and subsequently maintaining abstinence for  $16 \pm 12$  years by the time of interviews (19).

### DNA Preparation and Assessment of Allelic Frequencies

DNA was prepared from blood or cell lines (21–23) and carefully quantitated. DNAs from groups of 20 individuals of the same phenotype were combined. Pooled genotyping reduced costs and allowed us to assess high densities of genotypes in these subjects while providing no threat of loss of genetic confidentiality to these individual research volunteers. Hybridization probes were prepared with precautions to avoid contamination, as previously described [500k array set; Affymetrix, Santa Clara, CA, USA (20)]. A total of 150 ng of pooled DNA was digested by use of *StyI* or *NspI*, ligated to appropriate adaptors,

and amplified with a GeneAmp polymerase chain reaction (PCR) system 9700 (Applied Biosystems, Foster City, CA, USA) with 3 min at  $94^{\circ}\text{C}$ ; 30 cycles of 30 s at  $94^{\circ}\text{C}$ , 45 s at  $60^{\circ}\text{C}$ , 15 s at  $68^{\circ}\text{C}$ ; and a final 7-min  $68^{\circ}\text{C}$  extension. PCR products were purified (MinElute™ 96 UF kits; Qiagen, Valencia, CA, USA) and quantitated, then 40  $\mu\text{g}$  of PCR product was digested for 35 min at  $37^{\circ}\text{C}$  with 0.04 unit/ $\mu\text{L}$  DNase I to produce 30–100 bp fragments, which were end-labeled using terminal deoxynucleotidyl transferase and biotinylated dideoxynucleotides and hybridized to the appropriate 500k array (*Sty I* or *Nsp I* arrays) (Affymetrix). Arrays were stained and washed as described (Affymetrix Genechip Mapping Assay Manual) by use of immunopure streptavidin (Pierce, Milwaukee, WI, USA), biotinylated antistreptavidin antibody (Vector Labs, Burlingame, CA, USA), and R-phycoerythrin streptavidin (Molecular Probes, Eugene, OR, USA). Arrays were scanned and fluorescence intensities quantitated using an Affymetrix array scanner as described previously (20).

Chromosomal positions for each SNP were sought using Build 36.1 (National Center for Biotechnology Information) and NETAFFYX (Affymetrix) data. Allele frequencies for each SNP in each DNA pool were assessed based on hybridization-intensity signals from four arrays, allowing assessment of hybridization to the 12 “perfect-match” cells on each array that are complementary to the PCR products from alleles “A” and “B” for each diallelic SNP on sense and antisense strands. We eliminated SNPs on sex chromosomes to provide greater power based on inclusion of both males and females, and SNPs whose chromosomal positions could not be adequately determined.

Each array was analyzed as described (20), with background values subtracted, normalization to the highest values noted on the array, averaging of the hybridization intensities from the array cells that corresponded to the perfect-match A and B cells, calculation of A/B

ratios by dividing average normalized A values by average normalized B values, arctangent transformations to aid combination of data from arrays hybridized and scanned on different days, and determination of the average arctan value for each SNP from the 4 replicate arrays. A *t* statistic for the differences between nicotine abusers and controls was generated as described (20) for each SNP. We focused on SNPs that displayed *t* statistics with  $P < 0.05$  to  $P < 0.01$  for differences between nicotine-dependent versus control individuals who never smoked more than 101 cigarettes and nicotine-dependent current smokers versus individuals who were successful in maintaining abstinence. We sought evidence for clustering of these SNPs by focusing on chromosomal regions in which at least 3 to 4 of these outlier SNPs, assessed by at least two array types, lay within 25 kb of each other. We termed these clustered, nominally positive SNPs “clustered-positive SNPs,” and focused our analyses on regions containing these clustered-positive SNPs. We used different criteria for *P* values and numbers of SNPs per cluster in these different analyses so that the methods would be as identical as possible to those used for previously reported analyses of the appropriate comparison Uhl *et al.*, Liu/Dr̄gon *et al.*, and Beirut *et al.* datasets.

To seek additional support for the clusters of nominally positive SNPs from the current dataset, we sought convergence between data from these clustered nominally positive SNPs and clustered nominally positive SNPs, determined in the same way, from the datasets of Beirut *et al.* for 38,000 SNP GWA studies of light versus heavy smokers (15), Uhl *et al.* for 500–600,000 SNP GWA studies of individuals who were successful versus unsuccessful in quitting (14), and Liu/Dr̄gon *et al.* for 600,000–1 million SNP GWA studies of dependence on illegal addictive substances (20). To provide insights into some of the genes likely to harbor variants that contribute to individual differences in addiction vulnerability, we sought candidate genes that were identi-

fied by overlapping clusters of positive SNPs from each of these samples.

We used 10,000 Monte Carlo simulation trials to compare observed results to those expected by chance, as previously described (20). For each trial, a randomly selected set of SNPs from the current dataset was assessed to see if it provided results equal to or greater than the results that we actually observed. The number of trials for which the randomly selected SNPs displayed (at least) the same features displayed by the observed results was then tallied to generate an empirical *P* value. These simulations were thus correct for the number of repeated comparisons made in these analyses, an important consideration in evaluating these GWA datasets.

To assess the power of our current approach, we used current sample sizes and standard deviations, the program PS v2.1.31 (24,25) and  $\alpha = 0.05$ . To provide controls for the possibilities that abuser-control differences observed herein were attributable to occult ethnic/racial allele frequency differences or noisy assays, we assessed, respectively, the overlap between the results obtained here and the SNPs that displayed the largest allele frequency differences between African-American versus European-American control individuals and the largest assay “noise.”

*All supplementary materials are available online at [www.molmed.org](http://www.molmed.org).*

## RESULTS

We assessed allele frequencies in multiple pools of DNA from smokers who smoked when assessed, former smokers who quit prior to assessment, and individuals who never smoked significant numbers of cigarettes. Modest variability occurred among replicate arrays that assessed the same pool. Standard error of the mean (SEM) values were 0.039. For variability among the different pools used to assess the same phenotype, SEM was 0.024. For the smoker versus nonsmoker comparison, these samples and these estimates of variance thus pro-

vided 0.99, 0.98, 0.87, and 0.54 power to detect allele frequency differences of 12.5%, 10%, 7.5%, and 5%, respectively. For the comparison of continuing smokers versus successful quitters, the samples provided 0.97, 0.89, 0.66, and 0.33 power to detect differences of the same sizes. The allele frequency differences used in these power calculations were comparable to the actual mean allele frequency differences for the nominally significant SNPs for smoker versus nonsmoker and continuing smokers versus successful quitters in these studies (4% and 7%, respectively).

### Smokers versus Nonsmokers

When we compared nicotine-dependent individuals from these samples to non-dependent individuals who smoked fewer than 101 cigarettes in their lives, 27,803 SNPs displayed allele frequency differences with nominal  $P < 0.05$ . Of these, 3563 nominally positive SNPs fell into 665 clusters of at least 4 SNPs in which each nominally positive SNP was separated by  $\leq 25$  kb from at least 1 other nominally positive SNP. In each of these clusters, we also required that SNPs assayed on both array types display nominally positive results. These results identified 289 genes (Table S1). None of the 10,000 Monte Carlo simulation trials identified clustering that was this significant (thus  $P < 0.0001$ ). For the 5570 SNPs that fell into 1334 clusters of at least 3 SNPs, Monte Carlo *P* values were  $< 0.0001$ .

### Current Smokers versus Quitters

We also observed significant clustering of nominally positive SNPs when we compared current smokers with those who had successfully quit. There were 5408 SNPs that displayed allele frequency differences with nominal  $P < 0.01$ ; 914 of these nominally positive SNPs fell into 239 clusters of at least 3 nominally positive SNPs separated from each other by  $< 100$  kb. In each of these clusters, we again required that SNPs assayed on both array types display nominally positive results. These results iden-

**Table 1.** Nicotine dependence genes.<sup>a</sup>

Gene	Gene description	Chromosome	Base pair	Clustered SNPs			
				Hamer	MNB		Beirut
					I	II	
<i>A2BP1</i>	Ataxin 2-binding protein 1	16	6009133	6	3		13
<i>C10orf11</i>	Chromosome 10 open reading frame 11	10	77212525	9		3	1
<i>CAPN13</i>	Calpain 13	2	30799141	4	3		2
<i>CDH13</i>	Cadherin 13	16	81218079	14	6	3	2
<i>CHL1</i>	Close homolog of L1	3	213650	2	7		3
<i>CNTN6</i>	Contactin 6	3	1109629	9	4		6
<i>CSMD1</i>	CUB and Sushi multiple domains 1	8	2782789	14	15	10	12
<i>CSMD2</i>	CUB and Sushi multiple domains 2	1	33752196	5	4		
<i>CSRP3</i>	Cysteine and glycine-rich protein 3	11	19160154	3		3	
<i>CTNNA3</i>	Catenin $\alpha$ 3	10	67349937	4	4	3	21
<i>CTNND2</i>	Catenin delta 2	5	11024952	4	3		1
<i>CXCL14</i>	Chemokine (C-X-C motif) ligand 14	5	134934274	2		4	
<i>DDEF1</i>	Development and differentiation enhancing factor 1	8	131133535	4	3		
<i>FHIT</i>	Fragile histidine triad gene	3	59710076	15	3	5	2
<i>FHOD3</i>	Formin homology 2 domain containing 3	18	32131700	8		3	
<i>FLJ45872</i>	FLJ45872 protein	8	138890869	4	3		
<i>GRIK2</i>	Glutamate receptor, ionotropic, kainate 2	6	101953675	4		3	1
<i>GRM3</i>	Glutamate receptor, metabotropic 3	7	86111166	4	3		1
<i>HPSE2</i>	Heparanase 2	10	100208867	9	3		1
<i>LRP1B</i>	Low density lipoprotein-related protein 1B	2	140705466	4	5		4
<i>LRRN6C</i>	Leucine rich repeat neuronal 6C	9	27938528	7	4		18
<i>NME7</i>	Nucleoside-diphosphate kinase	1	167368393	4		14	
<i>PARD3B</i>	Par-3 partitioning defective 3 homolog B	2	205118761	4	3		
<i>PBX1</i>	Pre-B-cell leukemia transcription factor 1	1	162795561	4		5	
<i>PDE1C</i>	Phosphodiesterase 1C	7	31795772	9	3	3	
<i>PTPRM</i>	Receptor type protein tyrosine phosphatase M	18	7557817	5	3		
<i>RAD51L1</i>	RAD51-like 1	14	67356262	5		3	
<i>RYR3</i>	Ryanodine receptor 3	15	31390469	5	5		3
<i>SGCZ</i>	Sarcoglycan zeta	8	13991744	6	7		
<i>THSD4</i>	Thrombospondin type I dom 4	15	69220842	4	–	3	

<sup>a</sup>Genes that contain overlapping clusters of nominally positive SNPs in comparisons of nicotine-dependent individuals versus individuals who report use of <100 cigarettes in their lives (Hamer samples) as well as Liu/Drgon *et al.* and/or Beirut *et al.* comparisons. Columns list gene description, chromosome, base pairs for gene's start, number of nominally positive SNPs that lay in clusters within the gene's exons and in 10-kb genomic flanking regions, and the numbers of clustered, nominally positive SNPs that fall within overlapping clusters within the same gene from Liu/Drgon *et al.* The 600k (MNB I) and 1M (MNB II) data from the National Institute on Drug Abuse research volunteers dependent on at least one illegal substance and from the Beirut *et al.* 38k dataset (15) are listed in the last three columns. All of the genes that are identified by clustered nominally positive data from the current dataset are presented in Table S1.

tified 67 genes (Table S2). Monte Carlo *P* values for this degree of clustering were *P* < 0.0001.

#### Overlap with Beirut *et al.* and Liu/Drgon *et al.* Datasets

These data for clustered, nominally positive SNPs from the current dataset revealed significant overlap with genes identified by other relevant datasets. The clusters of 4 nominally positive SNPs identified by comparisons be-

tween dependent and nondependent subjects provided highly significant overlap with the subset of 38,000 SNPs that were identified as nominally significant by Beirut *et al.* in comparisons of dependent smokers to nondependent smokers (Monte Carlo *P* < 0.0001). Overlaps with genes identified in 600,000–1 million GWA for dependence on at least one illegal substance versus ethnically matched control individuals provided *P* values that reached statisti-

cal significance (*P* = 0.047). These overlaps identified 30 genes (Table 1).

#### Overlap with the Uhl *et al.* Dataset

The clusters of three nominally positive SNPs from the data that compare current to former smokers also overlap significantly with the clustered, nominally positive results from 2 of the 3 recently reported samples of individuals who were successful versus those who were unsuccessful in quitting smoking in

clinical trials. Comparisons between these data and results reported for samples from Rose and colleagues and from Niarua, David, and colleagues [samples II and III in (14)] each displayed significant Monte Carlo  $P$  values  $\leq 0.0001$ . Comparison with similarly analyzed samples from Lerman and coworkers (14), however, provided only a trend toward significance ( $P = 0.10$ ). Genes identified by clusters of nominally positive SNPs in current and prior studies of ability to successfully quit smoking include ataxin 2-binding protein 1 (3 clustered nominally positive SNPs in the current dataset); CUB and Sushi multiple domains (10 SNPs), Down syndrome cell adhesion molecule (3 SNPs), protocadherin 15 (3 SNPs), and the retinoic acid receptor  $\beta$  (3 SNPs) (Table 2).

### Possible Alternative Explanations for Observed Results

We would anticipate the observed, highly significant clustering of SNPs that display nominally positive results if many of these reproducibly positive SNPs lay near and were in linkage disequilibrium with functional allelic variants that distinguished substance-dependent subjects from control subjects. We would not anticipate this degree of clustering if the results were due to chance. The Monte Carlo  $P$  values noted here are thus likely to receive contributions from both the extent of linkage disequilibrium among the clustered, nominally positive SNPs and the extent of linkage disequilibrium between these SNPs and the functional haplotype(s) that lead to the association with substance dependence.

Neither control for occult stratification nor assay variability provide convincing alternative explanations for most of the data obtained here. We examined the overlap between the 3563 and 914 clustered positive SNPs from nicotine-dependence and quit-success analyses, respectively, and the 2.5% of the SNPs for which the noise in validating studies was highest as well as the 2.5% of SNPs that displayed the largest differences be-

**Table 2.** Quit success genes.<sup>a</sup>

Gene	Gene description	Chromosome	Base pair	Clustered SNPs
<i>A2BP1</i>	Ataxin 2-binding protein 1	16	6009133	3
<i>CSMD1</i>	CUB and Sushi multiple domains 1	8	2782789	10
<i>DSCAM</i>	Down syndrome cell adhesion molecule	21	40306213	3
<i>PCDH15</i>	Protocadherin 15	10	55250866	3
<i>RARB</i>	Retinoic acid receptor $\beta$	3	25444758	3

<sup>a</sup>Genes that contain overlapping clusters of nominally positive SNPs in comparisons of nicotine-dependent individuals who reported continuing smoking when interviewed versus those who reported prolonged abstinence and clusters of nominally positive SNPs from studies of at least two samples of successful versus unsuccessful quitters in clinical trials recently reported by Uhl *et al.* (14). Columns list gene description, chromosome, base pairs for gene's start, and number of nominally positive SNPs that lay in clusters within the gene's exons and in 10-kb genomic flanking regions in the current dataset. All genes identified by clustered, nominally positive data from the current dataset are presented in Table S2.

tween African-American versus European-American control individuals in Baltimore. We found 57 and 90, respectively, versus 89 and 89 expected by chance for dependence analyses. We found 23 and 15, respectively, versus 29 and 29 expected by chance for the quit-success analyses.

### DISCUSSION

The current results provide (a) independent support for GWA results from larger samples of individuals studied for other substance-dependence phenotypes, (b) independent support for GWA results for studies of smoking cessation, and (c) a control for one of the potential confounding features of previously reported smoking cessation GWA samples. The possibility that genetic results from members of any sample of research volunteers might not represent the genetics of members of the general population is ever present. In studies of molecular genetics in addiction, however, features that are both heritable and differentially present in substance-dependent individuals might, *a priori*, be considered to be especially likely to provide confounding influences (26).

Our present observations that findings from a community-based sample of re-

search volunteers converge with findings from two larger research-volunteer samples that were ascertained in different fashions are thus reassuring. Our findings also support the idea that many of the molecular genetic findings that have been previously reported are not simply attributable to the methods used for ascertainment and definition of "cases" and "controls." It is important to note, however, that this overall conclusion does not exclude sampling-related contributions related to the observations for some particular genes.

The observations comparing individuals who have smoked less than 101 cigarettes with a group of individuals who are currently nicotine dependent or were nicotine dependent in the past fit with many, but not all, aspects of the GWA Beirut *et al.* dataset (15). The overall fit between these datasets provides mutual support for a number of the genes identified herein. However, we identified no positive associations for SNPs in the chromosome 15 nicotinic acetylcholine receptor gene cluster. This gene cluster was previously identified by comparisons of heavy smokers with FTND-defined dependence versus smokers who were not nicotine dependent by FTND criteria but who were required to

have smoked more than 100 cigarettes in their lives. Almost 25% of the members of this “control” comparison group thus displayed nicotine dependence based on DSM criteria. The current results are consistent with those predicted from recent reports from Thorgerison and colleagues, in which SNP allele frequencies in the nicotine-receptor chromosome 15 acetylcholine-receptor gene cluster were similar in nonsmokers and heavy smokers (16). The significant differences that emerged in their data when both groups were compared with individuals who smoked more than 100 cigarettes in their lives but did not progress to heavy smoking and FTND nicotine dependence fit with the present data. Nondependent individuals who smoked at low and steady levels would fit into neither the nicotine-dependent nor the control group in our current study.

A number of limitations must be kept in mind in considering the present results. The preplanned approach used here demands that multiple nominally positive SNPs from each sample tag the same genomic region within a gene. Requirements that nominally positive SNPs from the current dataset come from each of the two 500k array types add a technical control. Monte Carlo approaches that do not require specification of underlying distributions can readily judge the degree to which all of the observations made here could be attributable to chance. Nevertheless, no unanimous criteria exist for declaring “replication” or “convergence” for GWA studies, a consideration worth considering in evaluating the current results. Another limitation is that these NIH research volunteer samples are of modest size. Power calculations that document the modest power in European-American samples reveal even more modest power for the smaller numbers of individuals of other racial/ethnic ancestries. We have thus not analyzed samples with non-European ancestries. This modest power limits interpretation of negative data and restricts inferences about genes that are not identified in

these samples. Although there is significant confidence in the overall set of convergent positive results reported here, tests for the significance of individual genes provide much more modest levels of statistical assurance.

Our focus on data from autosomal chromosomes allowed us to combine data from male and female subjects but at the cost of missing potentially important contributions from sex chromosomes. Our pooling approach provided excellent correlations between individually genotyped and pooled allele frequency assessments in validation experiments and allowed us to use these samples without adding additional confidentiality burdens. Nevertheless, estimates of allele frequencies based on pooled data represent approximations of “true” allele frequency differences that might be determined by error-free individual genotyping of each participant.

There is no indication that the overall positive results reported here are based on the SNPs whose assays provide more noise, and no indication that occult stratification on racial/ethnic lines contributed overall to the results that we obtain here. However, we cannot totally exclude contributions of occult stratification that cannot be detected by these overall screens to findings in specific genes. The convergence of data from smoking GWA with data derived from studies of individuals with addictions to substances in several different pharmacological classes supports the idea that many allelic variants enhance vulnerability to many addictions; the more modest statistical significance of this overlap fits with the idea that there is also likely to be some nicotine-selective genetics as well. We have no convincing explanation for the strong fit of our current quit-success molecular genetic results with data from 2 of the 3 previously reported samples but not with the third sample.

Our focus is on identification of genes. Although associations away from annotated genes can also provide interesting results, the genes that we identify in the

present work provide a number of interesting views of addiction. These data reinforce our observations that many of these genes are likely to contribute to brain differences that are reflected in the mnemonic aspects of addiction, and that some of them also provide tempting targets for antiaddiction therapeutics. We discuss these ideas in more detail elsewhere (26).

The functional classes into which the genes identified here are interesting, and too complex to be described in detail here. We remain impressed by the identification of a number of genes whose products are involved in cell adhesion, cell-to-cell communication processes that are likely to be key to proper establishment and plasticity of neuronal connections. *CUB and Sushi multiple domains 1 (CSMD1)*, *Down syndrome cell adhesion molecule (DSCAM)*, and *protocadherin 15 (PCDH15)* are 3 of the 5 “quit success” genes identified in Table 2. Each of these 14 of the 30 genes identified in Table 1 encodes a protein implicated in cell adhesion and/or extracellular matrix activities that are crucial to synapse formation and maintenance: *cadherin 13 (CDH13)*, *close homolog of L1 (CHL1)*, *contactin 6 (CNTN6)*, *CUB and Sushi multiple domains 1 and 2 (CSMD1, CSMD2)*, *catenin  $\alpha$  3 (CTNNA3)*, *catenin delta 2 (CTNND2)*, *fragile histidine triad gene (FHIT)*, *heparanase 2 (HPSE2)*, *low density lipoprotein-related protein 1B (LRP1B)*, *leucine rich repeat neuronal 6C (LRRN6C)*, *protein tyrosine phosphatase, receptor type, M (PTPRM)*, *sarcoglycan zeta (SGCZ)*, and *thrombospondin type 1 domain 4 (THSD4)*.

The significant convergence between the “quit success” molecular genetics identified here and data from 2 of the 3 currently reported samples supports the idea that genetic contributions to the ability of individuals to stop smoking in a community-based, often with modest pharmacological and behavioral support, are likely to overlap significantly with genetic contributions to the ability to quit smoking in research settings that frequently manifest substantially greater levels of pharmacological and behavioral support. These findings are thus reassuring.

ing for future attempts to use genotype-based data to match smokers with the types and/or intensities of treatments that provide the best and most cost-effective smoking cessation opportunities in community settings. Our findings also fit with the data from twin studies, because none of the twin studies of smoking cessation success was carried out in a clinical trial setting (10,11).

It appears likely that investigations of initial molecular genetic influences on therapeutic responses in other diseases may well be carried out by use of DNA from participants in clinical trials, because the intensive patient/subject monitoring that is often built into such trials provides reassurance that clinical outcomes are well documented in these datasets. However, the greatest utility of the information from genetic studies of clinical trial participants will come when there is additional data that can assure us that the results can be generalized to more real-world settings and to individuals who would be less likely to be enrolled in a clinical trial. The findings presented here thus support ongoing processes for comparing GWA datasets from more intensively studied and meticulously selected individuals with those from community-based samples. For addictions, as for many complex disorders, such data provide an increasingly rich basis for improved understanding and for development of personalized prevention and treatment strategies.

#### ACKNOWLEDGMENTS

We are grateful for thoughtful advice and discussion from Dr. J Rose and for fruitful collaborations with Drs. Rose, Lerman, Niaura, and David. This research was supported financially by the NIH Intramural Research Programs, National Institute on Drug Abuse, National Institute of Mental Health, and National Cancer Institute, Department of Health and Social Services.

#### DISCLOSURE

We declare that the authors have no competing interests as defined by *Molec-*

*ular Medicine*, or other interests that might be perceived to influence the results and discussion reported in this paper.

#### REFERENCES

- Uhl GR, Elmer GI, Labuda MC, Pickens RW. (1995) Genetic influences in drug abuse. In: Gloom FE, Kupfer DJ (eds.) *Psychopharmacology: The Fourth Generation of Progress*. New York: Raven Press. pp. 1793–2783.
- Tsuang MT, et al. (1998) Co-occurrence of abuse of different drugs in men: the role of drug-specific and shared vulnerabilities. *Arch. Gen. Psychiatry* 55:967–72.
- Karkowski LM, Prescott CA, Kendler KS. (2000) Multivariate assessment of factors influencing illicit substance use in twins from female-female pairs. *Am. J. Med. Genet.* 96:665–70.
- True WR, et al. (1999) Interrelationship of genetic and environmental influences on conduct disorder and alcohol and marijuana dependence symptoms. *Am. J. Med. Genet.* 88:391–7.
- Kendler KS, Karkowski LM, Neale MC, Prescott CA. (2000) Illicit psychoactive substance use, heavy use, abuse, and dependence in a US population-based sample of male twins. *Arch. Gen. Psychiatry* 57:261–9.
- Fagerstrom KO. (1978) Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addict. Behav.* 3:235–41.
- Fagerstrom KO, Schneider NG. (1989) Measuring nicotine dependence: a review of the Fagerstrom Tolerance Questionnaire. *J. Behav. Med.* 12:159–82.
- Pomerleau CS, Carton SM, Lutzke ML, Flessland KA, Pomerleau OF. (1994) Reliability of the Fagerstrom Tolerance Questionnaire and the Fagerstrom Test for Nicotine Dependence. *Addict. Behav.* 19:33–9.
- Pomerleau CS, Majchrzak MJ, Pomerleau OF. (1989) Nicotine dependence and the Fagerstrom Tolerance Questionnaire: a brief review. *J. Subst. Abuse* 1:471–7.
- Broms U, Silventoinen K, Madden PA, Heath AC, Kaprio J. (2006) Genetic architecture of smoking behavior: a study of Finnish adult twins. *Twin Res. Hum. Genet.* 9:64–72.
- Lessov CN, et al. (2004) Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychol. Med.* 34:865–79.
- Uhl GR, et al. (2008) Genome-wide association for methamphetamine dependence: convergent results from 2 samples. *Arch. Gen. Psychiatry* 65:345–55.
- Uhl GR, et al. (2007) Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs. *BMC Genet.* 8:10.
- Uhl GR, et al. (2008) Molecular genetics of successful smoking cessation: convergent genome-wide association study results. *Arch. Gen. Psychiatry* 65:683–93.
- Bierut LJ, et al. (2007) Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum. Mol. Genet.* 16:24–35.
- Thorgeirsson TE, et al. (2008) A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 452:638–42.
- Amos CI, et al. (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.* 40:616–22.
- Saccone SF, et al. (2007) Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum. Mol. Genet.* 16:36–49.
- Lueders KK, et al. (2002) Genetic and functional analysis of single nucleotide polymorphisms in the beta2-neuronal nicotinic acetylcholine receptor gene (CHRN2). *Nicotine Tob. Res.* 4:115–25.
- Liu QR, et al. (2006) Addiction molecular genetics: 639,401 SNP whole genome association identifies many “cell adhesion” genes. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 141B:918–25.
- Uhl GR, Liu QR, Walther D, Hess J, Naiman D. (2001) Polysubstance abuse-vulnerability genes: genome scans for association, using 1,004 subjects and 1,494 single-nucleotide polymorphisms. *Am. J. Hum. Genet.* 69:1290–300.
- Smith SS, et al. (1992) Genetic vulnerability to drug abuse. The D2 dopamine receptor Taq I B1 restriction fragment length polymorphism appears more frequently in polysubstance abusers. *Arch. Gen. Psychiatry* 49:723–7.
- Persico AM, Bird G, Gabbay FH, Uhl GR. (1996) D2 dopamine receptor gene TaqI A1 and B1 restriction fragment length polymorphisms: enhanced frequencies in psychostimulant-preferring polysubstance abusers. *Biol. Psychiatry* 40:776–84.
- Dupont WD, Plummer WD Jr. (1990) Power and sample size calculations. A review and computer program. *Control Clin. Trials* 11:116–28.
- Dupont WD, Plummer WD Jr. (1998) Power and sample size calculations for studies involving linear regression. *Control Clin. Trials* 19:589–601.
- Uhl GR, et al. (2008) Molecular genetics of addiction and related heritable phenotypes: genome-wide association approaches identify “connectivity constellation” and drug target genes with pleiotropic effects. *Ann. N. Y. Acad. Sci.* 1141:318–81.