

# Resequencing of Nicotinic Acetylcholine Receptor Genes and Association of Common and Rare Variants with the Fagerström Test for Nicotine Dependence

Jennifer Wessel<sup>1,4</sup>, Sarah M McDonald<sup>1</sup>, David A Hinds<sup>2</sup>, Renee P Stokowski<sup>2</sup>, Harold S Javitz<sup>1</sup>, Michael Kennemer<sup>2</sup>, Ruth Krasnow<sup>1</sup>, William Dirks<sup>2</sup>, Jill Hardin<sup>1</sup>, Steven J Pitts<sup>2</sup>, Martha Michel<sup>1</sup>, Lisa Jack<sup>1</sup>, Dennis G Ballinger<sup>2</sup>, Jennifer B McClure<sup>3</sup>, Gary E Swan<sup>1</sup> and Andrew W Bergen<sup>\*,1</sup>

<sup>1</sup>Center for Health Sciences, SRI International, Menlo Park, CA, USA; <sup>2</sup>Perlegen Sciences, Mountain View, CA, USA; <sup>3</sup>Group Health Research Institute, Seattle, WA, USA; <sup>4</sup>Department of Public Health, Indiana University School of Medicine, Indianapolis, IN, USA

Common single-nucleotide polymorphisms (SNPs) at nicotinic acetylcholine receptor (nAChR) subunit genes have previously been associated with measures of nicotine dependence. We investigated the contribution of common SNPs and rare single-nucleotide variants (SNVs) in nAChR genes to Fagerström test for nicotine dependence (FTND) scores in treatment-seeking smokers. Exons of 10 genes were resequenced with next-generation sequencing technology in 448 European-American participants of a smoking cessation trial, and *CHRNA2* and *CHRNA4* were resequenced by Sanger technology to improve sequence coverage. A total of 214 SNP/SNVs were identified, of which 19.2% were excluded from analyses because of reduced completion rate, 73.9% had minor allele frequencies <5%, and 48.1% were novel relative to dbSNP build 129. We tested associations of 173 SNP/SNVs with the FTND score using data obtained from 430 individuals (18 were excluded because of reduced completion rate) using linear regression for common, the cohort allelic sum test and the weighted sum statistic for rare, and the multivariate distance matrix regression method for both common and rare SNP/SNVs. Association testing with common SNPs with adjustment for correlated tests within each gene identified a significant association with two *CHRNA2* SNPs, eg, the minor allele of rs2072660 increased the mean FTND score by 0.6 Units ( $P=0.01$ ). We observed a significant evidence for association with the FTND score of common and rare SNP/SNVs at *CHRNA5* and *CHRNA2*, and of rare SNVs at *CHRNA4*. Both common and/or rare SNP/SNVs from multiple nAChR subunit genes are associated with the FTND score in this sample of treatment-seeking smokers.

*Neuropsychopharmacology* (2010) **35**, 2392–2402; doi:10.1038/npp.2010.120; published online 25 August 2010

**Keywords:** Fagerström test for nicotine dependence; single-nucleotide polymorphism; candidate gene association scan; treatment-seeking smokers

## INTRODUCTION

Cigarette smoking is the primary risk factor for many chronic diseases (Bergen and Caporaso, 1999). Genetic epidemiological studies have estimated the heritability of nicotine dependence to range from 40 to 75% (Broms, 2007 no. 1225; Li, 2003 no. 39; Maes, 2004 no. 402; Vink, 2005 no. 480; Lessov, 2004 no. 472). Recent candidate gene association studies (Baker *et al*, 2009; Bergen *et al*, 2009; Bierut *et al*, 2008; Breitling *et al*, 2009a,b; Chen *et al*, 2008, 2009; Etter *et al*, 2009; Freathy *et al*, 2009; Greenbaum *et al*, 2006; Hoft *et al*, 2009; Keskitalo *et al*, 2009; Le Marchand *et al*, 2008; Li *et al*, 2008, 2009; Perkins *et al*, 2008a,b, 2009;

Philibert *et al*, 2009; Rigbi *et al*, 2008; Saccone *et al*, 2009a,b; Sherva *et al*, 2008; Spitz *et al*, 2008; Stevens *et al*, 2008; Wang *et al*, 2009; Weiss *et al*, 2008), and genome-wide association scans (GWASs) (Drgon *et al*, 2009a,b; Liu *et al*, 2010, 2009; TAG, 2010; Thorgeirsson *et al*, 2008, 2010), with many published candidate gene and GWAS studies through 2009 referenced by Wang and Li (2010), have searched for, and, at varying levels of significance, identified, common variants associated with measures of response to tobacco, tobacco consumption, nicotine dependence, nicotine metabolism, or smoking cessation. However, most of the attributable fraction assumed to be due to heredity cannot be explained by these common single-nucleotide polymorphisms (SNPs), suggesting that rare variants in candidate genes, ie, those with minor allele frequencies (MAFs) substantially <5%, may also contribute to complex disease (Bodmer and Bonilla, 2008; Frazer *et al*, 2009; Schork *et al*, 2009). In contrast to common SNP associations, which typically exhibit odds ratios in case-control studies

\*Correspondence: Dr AW Bergen, Molecular Genetics Program, Center for Health Sciences, SRI International, 333 Ravenswood Avenue, SRI International, Menlo Park, CA 94025, USA, Tel: +1 650 859 4618, Fax: +1 650 859 5099, E-mail: andrew.bergen@sri.com  
Received 12 April 2010; revised 21 June 2010; accepted 12 July 2010

substantially less than two, rare variants at the genetic loci known to contribute to specific diseases exhibit substantially greater odds ratios, leading to the suggestion that despite their reduced allele frequency, rare variants might account for a substantial minor fraction of the population attributable risk due to genetic variation (Bodmer and Bonilla, 2008). Recent technological advances in high-throughput, next-generation sequencing enable direct sequencing of candidate genes, which is the optimal way to identify rare variants (Harismendy *et al*, 2009). A number of methods for identifying associations between rare variants and common diseases have been developed (Li and Leal, 2008; Madsen and Browning, 2009; Schork *et al*, 2008), which share the approach of combining information obtained from multiple variants.

In light of recent evidence from GWAS and candidate gene results, we chose 10 presynaptic neuronal nicotinic acetylcholine receptor (nAChR) subunit genes (namely *CHRNA2*, *CHRNA3*, *CHRNA4*, *CHRNA5*, *CHRNA6*, *CHRNA7*, *CHRNA1*, *CHRNA2*, *CHRNA3*, and *CHRNA4*) for resequencing to identify both common and rare variants for association analyses to measures of smoking behavior collected in a sample of treatment-seeking smokers involved in a randomized behavioral and pharmacological intervention for smoking cessation (Swan *et al*, 2010). Collectively, these genes code for receptors that bind the exogenous ligand nicotine, responsible for nicotine addiction (Changeux, 2009; Dani and Bertrand, 2007; Stolerman and Jarvis, 1995). On the basis of recent studies in the literature (Saccone *et al*, 2009a; Saccone *et al*, 2007), we hypothesized that common variation in these 10 nAChR subunit genes might contribute to variation in a commonly assessed measure of nicotine dependence, the Fagerström test for nicotine dependence (FTND) (Heatherton *et al*, 1991). In addition, on the basis of the emerging appreciation for the potential of rare variants to contribute to common disease (the common disease:rare variant hypothesis), we hypothesized that rare variation at these genes might also contribute to variation in this measure of nicotine dependence in this same sample. We employed a recently developed method and a traditional method (Margulies *et al*, 2005; Sanger *et al*, 1977) of resequencing to identify both common and rare sequence variations, and several recently developed methods for analysis of sequence variation that assesses the association between common and/or rare variants and phenotype combining information obtained from multiple variants to evaluate these hypotheses (Li and Leal, 2008; Madsen and Browning, 2009; Schork *et al*, 2008). Herein, we assess the contribution of common and/or rare sequence variation identified in 10 nAChR subunit genes in a sample of treatment-seeking smokers with the FTND score collected at the time of intake in a randomized smoking cessation trial.

## MATERIALS AND METHODS

### Population and Phenotype

Current smokers ( $\geq 10$  cigarettes per day over the past year,  $N = 1202$ ) were recruited from the Group Health (GH), a consumer-governed nonprofit health-care organization that serves  $> 630\,000$  residents of Washington and Idaho,

to participate in a smoking cessation randomized behavioral intervention combined with the FDA-approved course of varenicline. Recruitment and treatment methods for the Comprehensive Medication Program and Support Services (COMPASS) study, a randomized trial (NCT00301145) sponsored by the National Cancer Institute, have been described previously (Swan *et al*, 2010). Nicotine dependence was measured at baseline with the FTND. Scores on this scale range from 0 to 10 with a higher score indicating greater dependence; although the FTND has suboptimal psychometric properties, it continues to be used in part because of the amount of previous clinical, descriptive, and genetic research that has used it, and in part because of the modest associations observed with the clinical outcome of smoking cessation (Piper *et al*, 2006). DNA was extracted from saliva samples provided by COMPASS participants (Nishita *et al*, 2009), and DNA obtained from 448 individuals who self-identified as belonging to non-Hispanic White race/ethnicity, who had never used varenicline previously, and who had questionnaire data on smoking habits, was used for genetic analyses. All protocols were reviewed and approved by the Institutional Review Boards of the GH and SRI International.

### 454 Resequencing

We designed target regions (amplicons) to amplify nAChR subunit gene exons and flanking regions including splice sites and selected 5' and 3' untranslated regions (UTRs) and SNPs upstream and downstream of transcription units. Primer design, sample preparation, and sequencing followed the protocols supplied by Roche for the 454 Genome Sequencer System (Supplementary File 1). We used the Roche GS Reference Mapper alignment tool to align sequence reads to a target consisting of the amplified intervals from the reference genome (NCBI Build 36). We separately used BLAST (Altschul *et al*, 1990) to align each read against the entire reference genome, and removed reads with the best BLAST hits that did not match the GS Reference Mapper alignment to the target. Candidate single-nucleotide variants (SNVs) were identified by parsing the GS Reference Mapper alignments and identifying positions at which at least one individual had reads of the same alternate base in forward and reverse orientations. We then used a Bayesian algorithm similar to that used in the study by Craig *et al* (2008) to assign genotypes and quality scores to all sequenced individuals at each putative variant position based on read counts for the reference and alternate base. This approach assigns genotypes taking into account genotype frequencies and error rates estimated at each variant position. Genotypes that were less than twice as likely as the next most likely genotype were coded as missing. Positions that were substantially out of the Hardy-Weinberg equilibrium ( $P < 10^{-7}$ ) were excluded from analysis; the Hardy-Weinberg equilibrium test has power to detect only disequilibrium with common SNPs.

### Sample Quality Control

We excluded 18 DNA samples with overall call rates  $< 90\%$ . The amplified regions included 29 SNPs in the Phase II HapMap release, 24 that were polymorphic in the CEU

(Utah residents with Northern and Western European ancestry from the CEPH collection) panel. All of these SNPs were identified as polymorphic in the sequencing data, although four failed a quality control (QC) threshold as a result of fold coverage  $<10 \times$ . Allele frequencies for the remaining 25 were within 6% of HapMap CEU frequencies. We included 10 additional amplicons containing 1 QC SNP each and a single Y-chromosome amplicon as control amplicons. Each of the sequenced DNA samples was also genotyped with a custom high-density oligonucleotide array containing 312 SNPs that contained the 10 QC SNPs (Hinds *et al*, 2004). Principal components analysis of the SNP array data did not indicate the presence of population structure, and confirmed that all samples were of European ancestry. No samples were excluded from association analyses because of a high discordance rate between the genotypes for the 10 QC SNPs from the array and sequencing data sets (data not shown).

### Sanger Resequencing

Selected regions of *CHRNA2* and *CHRNA4* with low read depth ( $<10X$ ) from 454 resequencing were resequenced using Sanger sequencing methods (Sanger *et al*, 1977), primers as developed by Weiss *et al* (2008) or as optimized in this study, enzymatic methods for PCR as optimized in this study, and using BigDye Terminator v3.1 Cycle Sequencing protocols on a Applied Biosystems 3130xl Genetic Analyzer. Variant detection was performed using Mutation Surveyor v3.24 (SoftGenetics, LLC, Penn State, Pennsylvania), Reference Sequence (RefSeq) NM\_000748.2 for *CHRNA2* and NM\_000744.5 for *CHRNA4* and reference files generated in the Mutation Surveyor using the RefSeq and the 'generate synthetic chromatograms' function. See Supplementary File 1 for further details.

### Association Analyses

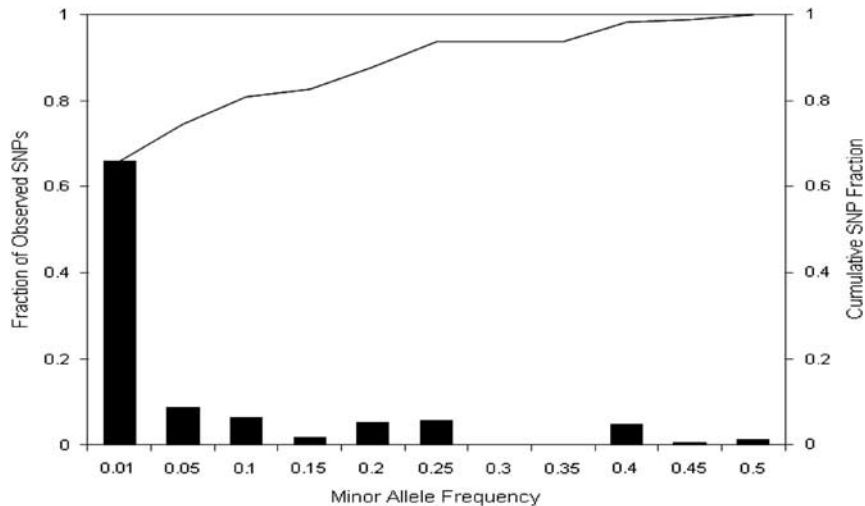
Common variants (MAF  $\geq 5\%$ ) were analyzed separately for association with normally distributed FTND using a generalized linear model. Genotypes were coded in additive (Add) or dominant (Dom) models. The significance of regression models is reported for each SNP and with adjustment for correlated tests ( $P_{ACT}$ ) (Conneely and Boehnke, 2007) for the most significantly associated SNP within each gene and model to a  $P_{ACT}$  value threshold of 0.10. For rare variants, gene-based association tests were performed by the cohort allelic sum test (CAST) and by the weighted sum statistic (WSS) (Madsen and Browning, 2009). CAST tests for association between FTND and counts of rare alleles were based on two fixed thresholds for 'rareness' (MAF  $<0.01$  and  $<0.05$ ). The WSS was used to test for association between FTND and weighted counts of minor alleles (MAF  $<0.05$ ), with an inverse relationship between weights and the MAF of minor alleles. Both tests are applied only to rare variants under the assumption that rare variants are more likely to be deleterious than common variants (Kryukov *et al*, 2009). Linear regression coefficients,  $P$ -values from likelihood ratio tests, and empirical  $P$ -values from permutation testing are reported. Multivariate distance-based matrix regression (MDMR) was also employed to test associations of common and rare

(MAF  $<0.05$ ) variants with FTND, with either identical-by-state allele sharing across individuals and variants in each gene, and with allele sharing weighted by the Lynch–Ritland calculation, with 100 000 permutations. The latter approach gives more weight to rare variants (Nievergelt *et al*, 2007; Schork *et al*, 2008; Wessel and Schork, 2006). For nAChR subunit genes that are clustered in the genome (*CHRNA3* and *CHRNA6* at chr8p11, and *CHRNA5*, *CHRNA3*, and *CHRNA4* at chr15q25.1), we also performed association analyses to evaluate all sequenced SNPs available in these genes together. When MDMR analyses with both common and rare SNPs identified a significant association, we performed and report results from two *post hoc* tests: analysis of rare SNPs alone and analysis of common SNPs alone. Power to detect effect sizes with parameters given,  $\alpha = 0.05$  and a two-sided test was estimated using Quanto software (Gauderman, 2002). Pairwise linkage disequilibrium (LD) ( $D'$  and  $r^2$ ) values were calculated from the COMPASS sequence data using Haploview (Barrett *et al*, 2005). Genes and SNPs are ordered in tables by the chromosome and coordinate, respectively.

## RESULTS

### nAChR Exome Sequencing and SNP Discovery

We attempted to resequence all coding exons, splice sites, and selected SNPs at 10 nAChR subunit genes comprising 18 kb of target sequence in 448 European ancestry treatment-seeking participants of the COMPASS clinical trial using 454 sequencing supplemented by Sanger sequencing for two genes (Supplementary File 1). For the 10 nAChR subunit genes sequenced by 454 technology, 66% of targeted base pairs in these amplicons had  $>10$  reads, with 3 genes accounting for most of the shortfall in target coverage (*CHRNA4*, 0%; *CHRNA2*, 22%, and *CHRNA7*, 39%, vs an average for the 7 other genes of 81%, please see Table 3 and Figure 1 in Supplementary File 1). A total of 167 unique SNPs were identified through 454 sequencing. Before analyses, we excluded 18 DNA samples and 24 SNPs that had call rates  $<90\%$  and 5 SNPs that were monomorphic in the remaining 430 samples. The mean (SD) Sanger amplicon resequencing read depth was 1.98 (0.20) and 1.88 (0.43) for *CHRNA2* and *CHRNA4*, respectively, and SNPs were scored using default base-calling parameters. Sanger resequencing identified 15 SNPs at *CHRNA2* and 32 SNPs at *CHRNA4* (Supplementary File 1 and Table 1). Six SNPs found in *CHRNA4* exon 5 and *CHRNA2* exon 5 by 454 resequencing were not found by Sanger resequencing and hence were excluded. One SNP, found by both methods, was unreliable in both methods, because of poor coverage in 454 resequencing and unclear reads in Sanger resequencing and hence was also excluded. Using the two methods, a total of 172 SNPs were identified in 430 samples (Supplementary Table 1). The observed distribution of MAFs in the total sample is strongly skewed to low MAF values (Figure 1 and Supplementary Table 1), in which 73.9% have MAF  $<5\%$  and 65.3% have MAF  $<1\%$ , with MAFs ranging from 0.1 to 49.2%. Percentages and types of variants identified (Supplementary Table 1) are 4.1% 5' of the 5' UTR, 4.1% 5' UTR, 51.4% coding (26.0% nonsynonymous, 25.4%



**Figure 1** Minor allele frequencies (x axis), proportion (left y axis) and cumulative fraction (right y axis) of nAChR SNPs identified through resequencing. Bars represent the percentage of SNPs within each category of minor allele frequency (MAF) with the figure under each bar being the maximum MAF within that category. The line represents the cumulative frequency of SNPs.

synonymous), 23.1% intronic, 14.5% 3' UTR, and 2.9% 3' of the 3' UTR.

### Treatment-Seeking Smoker Characteristics

The 430 individuals included in association analyses of nAChR subunit gene SNPs and the FTND score are mostly female (68%), with a mean (SD) age of 49.0 years (11.3) and of cigarettes smoked per day 20.4 (8.5). In this sample of 430 treatment-seeking smokers, gender and age were not significantly associated with FTND ( $P > 0.05$ ). The sample had mean (SD) FTND scores of 5.2 (2.1), indicating moderate addiction to tobacco (Heatherton *et al*, 1991). This mean FTND score is higher than population-based samples (Fagerstrom and Furberg, 2008), but typical of treatment-seeking smokers from this patient population (Swan *et al*, 2003; Swan *et al*, 2010) and other treatment-seeking smoker samples (Bergen *et al*, 2009).

### Common Variation Association Analyses

In all, 44 SNPs with  $MAF \geq 5\%$  were tested for association with FTND using two transmission models. Significant ( $P < 0.05$ ) unadjusted associations were found between *CHRNA1* (rs2302764 and rs7210231,  $D' = 1$ ,  $r^2 = 0.049$ ) and *CHRNA2* (rs2072660 and rs2072661,  $D' = 1$ ,  $r^2 = 0.918$ ) SNPs with FTND (Table 1). After adjustment for multiple correlated tests within each gene, significant associations remained between two *CHRNA2* SNPs located in the 3' UTR and the FTND (rs2072660,  $P_{ACT, Additive} = 0.013$ , and rs2072661,  $P_{ACT, Additive} = 0.009$ ). Individuals with one or two copies of the minor allele of these *CHRNA2* SNPs exhibit a 0.6 unit increase in the mean FTND score (for rs2072661, FTND mean (SD) = 5.5 (0.2) vs 4.9 (0.1)). After adjustment for correlated tests, the minor allele of the *CHRNA1* SNP rs7210231 was nonsignificantly ( $P_{ACT, Dominant} = 0.082$ ) associated with a decrease in the FTND score. In a *post hoc* test of mean FTND scores, individuals with the

rs7210231 heterozygote genotype exhibit significantly lower mean FTND scores than do individuals with homozygote genotypes (heterozygote mean (SE) = 4.70 (0.19), major allele homozygote = 5.77 (0.50), minor allele homozygote mean (SE) = 5.36 (0.13), with heterozygote to major homozygote  $P = 0.044$  and heterozygote to minor homozygote  $P = 0.004$ ).

### Common and Rare Variation Analyses

Association with rare SNPs within a gene and the FTND are reported in Table 2 for WSS ( $MAF < 0.05$ ) and CAST (two thresholds,  $MAF < 0.01$  and  $< 0.05$ ) analyses, and associations with common and rare ( $MAF < 0.05$ ) SNPs and FTND score are reported in Table 3 for MDMR analyses, along with *post hoc* MDMR tests. Significant ( $P < 0.05$ ) association with rare variation in *CHRNA4* and the FTND score was observed using the CAST tests, but not the WSS test (Table 2). Significant associations between common and rare variants and the FTND score were identified at *CHRNA2*, *CHRNA5*, and at the chr15q25.1 nAChR locus using the MDMR method (Table 3). When either common or rare variants at these genes were removed before association analysis, associations between rare or common SNPs at *CHRNA5* and at the chr15q25.1 nAChR locus and the FTND score, and those between rare *CHRNA2* SNPs and the FTND score, were no longer significant. Some association results that did not reach the threshold of statistical significance used herein, eg,  $P$ -values between 0.05 and 0.10, suggest that common variation at *CHRNA3* (Table 1), rare variation at *CHRNA2* and *CHRNA6* (Table 2), common and rare variations at *CHRNA2* (Table 3), and common and/or rare variation at *CHRNA1* (Tables 1 and 3) may also contribute to the FTND score in this sample of treatment-seeking smokers. We note that *CHRNA6* and *CHRNA7* had very limited numbers of variants available for analysis, and thus tests for association at these genes that combine variants will have low power.



**Table 1** Common nAChR Subunit Gene Variant Associations with the FTND Score

SNP ID	A1 <sup>a</sup>	A2 <sup>b</sup>	Gene	Type	MAF	Add P <sup>c</sup>	Dom P <sup>d</sup>
rs2280781	C	T	CHRNA2	5' UTR	0.101	0.221	0.280
rs4845378	G	T	CHRNA2	Intron	0.097	0.258	0.258
rs2072660	T	C	CHRNA2	3' UTR	0.244	<b>0.003</b>	<b>0.006</b>
rs2072661	G	A	CHRNA2	3' UTR	0.245	<b>0.002</b>	<b>0.004</b>
rs4292956	C	T	CHRNA2	3' UTR	0.066	0.472	0.476
rs2472553	G	A	CHRNA2	Non-syn	0.139	0.571	0.493
rs13277254	G	A	CHRNA3	Up	0.234	0.829	0.875
rs13280301	A	G	CHRNA3	Up	0.178	0.675	0.798
rs6474413	C	T	CHRNA3	Up	0.232	0.746	0.702
rs4950	G	A	CHRNA3	5' UTR	0.236	0.808	0.854
rs2304297	G	C	CHRNA6	3' UTR	0.244	0.557	0.985
ss161151744	C	T	CHRNA7	Syn	0.066	0.314	0.314
rs569207	C	T	CHRNA5	Intron	0.196	0.134	0.149
rs16969968	G	A	CHRNA5	Non-syn	0.367	0.433	0.322
rs615470	T	C	CHRNA5	3' UTR	0.381	0.330	0.329
rs8192482	C	T	CHRNA5	3' UTR	0.368	0.489	0.388
rs564585	A	G	CHRNA5	3' UTR	0.237	0.080	0.139
rs12899226	T	G	CHRNA3	Down	0.052	0.104	0.104
rs660652	A	G	CHRNA3	3' UTR	0.383	0.283	0.329
rs472054	A	G	CHRNA3	3' UTR	0.379	0.381	0.367
rs578776	A	G	CHRNA3	3' UTR	0.247	0.053	0.112
rs1051730	G	A	CHRNA3	Syn	0.367	0.455	0.359
rs3743075	T	C	CHRNA3	Syn	0.378	0.292	0.305
rs3743074	G	A	CHRNA3	Intron	0.378	0.336	0.324
rs8040868	T	C	CHRNA3	Syn	0.429	0.981	0.813
rs8192475	C	T	CHRNA3	Non-syn	0.050	0.094	0.094
rs12914008	G	A	CHRNA4	Non-syn	0.051	0.100	0.100
rs3813567	G	A	CHRNA4	Up	0.157	0.370	0.275
rs2302765	T	C	CHRNA1	Intron	0.159	0.729	1.000
rs12452047	A	G	CHRNA1	Intron	0.166	0.616	0.808
rs7210231	C	A	CHRNA1	Intron	0.199	0.124	<b>0.017</b>
rs2302761	C	T	CHRNA1	Intron	0.202	0.233	0.078
rs2302763	T	C	CHRNA1	Intron	0.164	0.679	0.889
rs2302764	T	C	CHRNA1	3' UTR	0.160	<b>0.049</b>	<b>0.045</b>
rs3827020	T	C	CHRNA4	Intron	0.143	0.521	0.865
rs45442394	G	A	CHRNA4	Intron	0.060	0.431	0.717
rs1044397	C	T	CHRNA4	Syn	0.492	0.305	0.608
rs1044396	G	A	CHRNA4	Syn	0.488	0.215	0.467
rs2229960	A	G	CHRNA4	Syn	0.057	0.870	0.758
rs2229959	C	A	CHRNA4	Syn	0.119	0.837	0.806
rs1044394	A	G	CHRNA4	Syn	0.067	0.784	0.678
rs6090384	T	C	CHRNA4	Intron	0.060	0.830	0.933
rs2273505	C	T	CHRNA4	Intron	0.062	0.740	0.725
rs2273506	G	A	CHRNA4	Syn	0.061	0.580	0.562

<sup>a</sup>A1 = allele 1.<sup>b</sup>A2 = allele 2.<sup>c</sup>Add = additive model.<sup>d</sup>Dom = dominant model.The bold values are for  $P < 0.05$ .

## DISCUSSION

We resequenced the exons of 10 highly relevant candidate genes for the study of nicotine dependence (Changeux,

**Table 2** Rare nAChR Subunit Gene Variant Associations with the FTND Score

	CAST <sub>β</sub> <sup>a</sup>	P <sub>1</sub> <sup>b</sup>	P <sub>2</sub> <sup>c</sup>	CAST <sub>β</sub> <sup>d</sup>	P <sub>1</sub>	P <sub>2</sub>	WSS <sub>β</sub> <sup>e</sup>	P <sub>1</sub>	P <sub>2</sub>
CHRNA2	-1.51	0.09	0.08	-1.51	0.09	0.08	-0.0067	0.09	0.10
CHRNA2	0.48	0.53	0.55	0.48	0.53	0.49	0.0007	0.48	0.53
CHRNA3	0.85	0.49	0.52	0.85	0.49	0.54	0.0038	0.49	0.59
CHRNA6	2.86	0.06	0.12	2.86	0.06	0.08	0.0127	0.06	0.08
chr8p11	1.66	0.08	0.07	1.66	0.08	0.08	0.0074	0.08	0.07
CHRNA7	1.35	0.37	0.33	1.35	0.37	0.35	0.0060	0.38	0.32
CHRNA5	0.10	0.93	0.94	0.61	0.19	0.23	-0.0005	0.54	0.60
CHRNA3	-2.17	0.15	0.10	-2.17	0.15	0.22	0.0022	0.15	0.19
CHRNA4	0.49	0.45	0.40	-0.13	0.75	0.67	0.0013	0.31	0.23
chr15q25.1	0.08	0.88	0.86	0.10	0.74	0.74	0.0002	0.78	0.77
CHRNA1	-1.17	0.28	0.33	-1.17	0.28	0.32	-0.0052	0.28	0.25
CHRNA4	-1.70	<b>0.04</b>	<b>0.04</b>	-1.70	<b>0.04</b>	<b>0.03</b>	-0.0011	0.45	0.43

<sup>a</sup>CAST-β, MAF < 0.01.<sup>b</sup>P<sub>1</sub> = P-value from standard F-test.<sup>c</sup>P<sub>2</sub> = P-value from permutation testing.<sup>d</sup>CAST-β, MAF < 0.05.<sup>e</sup>WSS-β.The bold values are for  $P < 0.05$ .

2009), using DNA extracted from saliva (Nishita *et al*, 2009), provided by treatment-seeking participants of a randomized smoking cessation trial (Swan *et al*, 2010), using a next-generation sequencing method (Margulies *et al*, 2005), together with the Sanger method (Sanger *et al*, 1977). We identified sequence variation and employed standard regression and specialized analysis methods (Madsen and Browning, 2009; Wessel and Schork, 2006) for analyzing association of this variation with the FTND score. The latter analysis methods group SNPs from a gene together into a summary variable to test for an excess of variants associated with the phenotype (Li and Leal, 2008). We selected methods to analyze only common (regression), only rare (CAST and WSS), and both common and rare (MDMR) variations at a gene to evaluate the possible influence of common and rare variations under the common disease:common variant and the common disease:rare variant hypotheses (Bodmer and Bonilla, 2008; Schork *et al*, 2009).

In the sample of 430 DNA samples obtained from treatment-seeking smokers and 173 nAChR subunit gene SNPs analyzed for association, we observed that common variants at *CHRNA2* and rare variants at *CHRNA4* were significantly associated with the FTND score, after correction for multiple correlated tests and by permutation testing within each gene, respectively. Employment of a method (allele sharing and weighted allele sharing MDMR with permutation testing) that assesses the contribution of common and rare variants combined, and *post hoc* analyses with rare or common variants alone, identified a significant association with variation in the FTND score at *CHRNA2*, *CHRNA5*, and at the chr15q25.1nAChR locus with both types of variation, and with the combination of common variants alone at *CHRNA2*.

There are at least two possible mechanisms underlying the latter findings described above, ie, the observation that the combination of common SNPs and rare variants, or of common SNPs alone, at a gene or gene region, are

**Table 3** Common and Rare nAChR Subunit Gene Variant Associations with the FTND Score

Gene	N SNPs	Allele sharing			Weighted allele sharing		
		Pseudo-F	P-value	% Variation	Pseudo-F	P-value	% Variation
<i>CHRNA2</i>	24	<b>7.57</b>	<b>0.002</b>	<b>0.017</b>	<b>4.53</b>	<b>0.019</b>	<b>0.011</b>
<i>CHRNA2</i> <sup>a</sup>	19	0.46	0.586	0.001	2.22	0.137	0.005
<i>CHRNA2</i> <sup>b</sup>	5	<b>7.51</b>	<b>0.003</b>	<b>0.018</b>	<b>5.48</b>	<b>0.010</b>	<b>0.013</b>
<i>CHRNA2</i>	11	0.69	0.503	0.002	23.23	0.095	0.052
<i>CHRNA3</i>	12	0.11	0.822	0.000	0.20	0.677	0.001
<i>CHRNA6</i>	3	0.63	0.436	0.002	4.24	0.361	0.010
chr8p11	15	0.09	0.838	0.000	0.20	0.702	0.001
<i>CHRNA7</i>	4	1.24	0.275	0.003	-12.02	0.814	-0.029
<i>CHRNA5</i>	15	1.27	0.299	0.003	<b>3.92</b>	<b>0.046</b>	<b>0.009</b>
<i>CHRNA5</i> <sup>a</sup>	10	-0.48	0.886	-0.001	0.55	0.556	0.001
<i>CHRNA5</i> <sup>b</sup>	5	1.85	0.177	0.005	2.64	0.124	0.007
<i>CHRNA3</i>	33	1.59	0.203	0.004	1.86	0.199	0.004
<i>CHRNA4</i>	15	1.16	0.343	0.003	0.12	0.807	0.000
chr15q25.1	63	1.40	0.249	0.003	<b>3.44</b>	<b>0.044</b>	<b>0.008</b>
chr15q25.1 <sup>a</sup>	47	0.911	0.543	0.002	-0.18	0.896	0.000
chr15q25.1 <sup>b</sup>	16	1.66	0.198	0.004	2.74	0.107	0.007
<i>CHRNA1</i>	25	1.67	0.206	0.004	2.84	0.064	0.007
<i>CHRNA4</i>	31	0.53	0.616	0.001	21.55	0.169	0.050

<sup>a</sup>Post hoc MDMR test performed with rare variants only.

<sup>b</sup>Post hoc MDMR test performed with common variants only.

The bold values are for  $P < 0.05$ .

associated with FTND, but that rare variants do not exhibit association. One possible mechanism is the interaction between common variants at the same gene or within the same gene region, which is the mechanism that has been elucidated by the studies by Li *et al* for the chr15q25.1 nAChR cluster, in which the interaction among *CHRNA3* SNPs was observed in association with measures of nicotine dependence (Li *et al*, 2009; Lou *et al*, 2008). Formal interaction analyses within genes or within multigene regions will be necessary to explore the possible role of this mechanism in the observed association of common and rare or common variants alone with the FTND score in the COMPASS sample. Another possible mechanism is the phenomenon of common SNPs acting as tagging SNPs for multiple rare causal variants that exist in the same LD region, ie, so-called synthetic associations (Dickson *et al*, 2010; Wang *et al*, 2010). When we removed common SNPs from the association test as we did when testing only rare variants in the MDMR analysis, we may have removed the only SNPs that have the power to detect the association of rare causal variants with the FTND score. It is well known that the chr15q25.1 nAChR locus exhibits extensive LD across a region of > 200 000 bp, and that the chr1q21.3 *CHRNA2* gene region exhibits strong LD extending ~150 000 bp in the HapMap CEU sample 3' from *CHRNA2* through *ADAR* to the 3' exons of *KCNN3*; thus conditions exist for synthetic associations to be present at the chr15q25.1 and chr1q21.3 nAChR loci. Resequencing of selected individuals based on a common SNP allele or haplotype content, followed by common and rare variant analyses, will be necessary to test the synthetic association hypothesis in these two regions.

Three other groups have resequenced nAChR subunit genes to identify variants for subsequent association analysis. The first group used DNA obtained from 192 European Americans (144 smokers and 48 nonsmokers sampled from low (0–4) and high (6–10) FTND strata) to resequence 10 genes and identified 262 SNPs (Weiss *et al*, 2008). They identified the chr15q25.1 nAChR locus to be significantly associated with the FTND strata in 144 smokers (72 high vs 72 low FTND strata) and genotyped 87 tag SNPs selected with MAF > 0.05 across the 10 genes in 3 larger cohorts (total  $N = 2827$ ), and a significant association between 6 chr15q25.1 nAChR locus SNPs with the FTND strata was observed among individuals initiating smoking at age  $\leq 16$  years. The size of the COMPASS sample is approximately one-sixth that of the sample used by Weiss *et al* to characterize the effect size of chr15q25.1 nAChR locus haplotypes. Replication of the Weiss *et al* finding will require the use of a larger sample of treatment-seeking smokers than available in the COMPASS sample. Sabatelli *et al* (2009) resequenced 3 nAChR subunit genes (namely *CHRNA3*, *CHRNA4*, and *CHRNA4*) in 245 sporadic amyotrophic lateral sclerosis (SALS) cases and 450 controls from Italy identifying 55 SNPs; significantly more missense variants in the SALS cases compared with controls were observed in the intracellular loop region of nAChRs ( $P = 0.0001$ ). In our analyses, we identified and evaluated 45 nonsynonymous variants at 10 nAChR subunit genes for association with the FTND score, most of which are rare (38 with MAF < 0.01 and 41 with MAF < 0.05). Using several methods to assess the association with rare variants in the COMPASS sample, we identified evidence for

a significant association with rare variants at *CHRNA4* with the FTND score. We observed a total of 21 rare (20 variants with  $MAF < 0.01$  and 1 variant with  $MAF < 0.05$ ) *CHRNA4* variants, with 5 nonsynonymous substitutions, including a nonsense variant, in the COMPASS sample analyzed. Further stratification of association testing of nAChR subunit gene variants in the COMPASS sample by variant type, nAChR structural region, and evaluation of function by *in vitro* expression and electrophysiology would be necessary to test the finding of Sabetelli *et al*. Rana *et al* (2009) resequenced the 3 chr15q25.1nAChR subunit genes in 80 individuals from 5 ethnicities identifying 63 variants; upon genotyping of 6 tag SNPs, significant associations in 370 European-American twins with *CHRNA3* SNPs rs3743075 and rs3743074 (in strong LD) were found with plasma catestatin and systolic blood pressure levels. These common *CHRNA3* SNPs were not associated with the FTND score in the COMPASS sample.

It is noteworthy that three recent meta-analyses of GWAS of smoking phenotypes in DNA samples obtained from population and clinically based data sets exceeding 140 000 individuals *in toto* have identified several common SNPs at the chr15q25.1 nAChR locus (rs55853698 and rs6495308 (Liu *et al*, 2010); rs1051730 and rs16969968 (TAG, 2010); and rs1051730 (Thorgeirsson *et al*, 2010)) associated with smoking quantity. In addition, one of these meta-analyses also identified common SNPs at the chr8p11nAChR locus (rs6474412 and rs13280604 (Thorgeirsson *et al*, 2010)) associated with smoking quantity. Of these six SNPs, our resequencing study identified only rs1051730 and rs16969968, and neither of these SNPs was statistically significantly associated with the FTND score in our sample. However, we did observe an association with multiple common SNPs at *CHRNA5* and at the chr15q25.1 nAChR locus using the MDMR approach, although none of the 16 common chr15q25.1 nAChR SNPs showed a nominally significant association ( $P < 0.05$ ) in individual regression models.

Our finding of a significant association of *CHRN2* common variants with a measure of nicotine dependence is supported by some previous candidate gene studies reporting significant associations with the FTND and related phenotypes in treatment-seeking, in laboratory, and in population and clinically based data sets distinct from the treatment-seeking smokers we studied (Swan *et al*, 2010). An analysis of 417 treatment-seeking smokers from a placebo-controlled randomized clinical trial of bupropion and placebo (Lerman *et al*, 2006) reported significant associations with *CHRN2* SNPs rs2072660 and rs2072661 and abstinence in a logistic model incorporating main effects and interaction with treatment at the end of treatment ( $P = 0.01$ ) and at 6 months ( $P = 0.0002$ ) (Conti *et al*, 2008). A short-term test of nicotine *vs* placebo patch effects in 156 smokers reported a significant association between rs2072661 and the number of abstinent days (Perkins *et al*, 2009). In an analysis of *CHRNA4* and *CHRN2* SNPs in a sample of 1068 older adolescents, rs2072660 was reported to be significantly associated with sensitivity to tobacco (subjective effects of 'nauseous' or 'dizzy') (Ehringer *et al*, 2007).

However, candidate gene studies evaluating the association of *CHRN2* SNPs rs2072660 and/or rs2072661 with

smoking status and measures of nicotine dependence have more often failed to identify significant associations. These included studies recruiting individuals, twins, or families from the community, including 2037 individuals from 602 multiplex heavy-smoking pedigrees (Li *et al*, 2005), 1929 ever-smoking unrelated individuals (Saccone *et al*, 2009a; Saccone *et al*, 2007), 872 unrelated members of a population-based twin sample (Silverman *et al*, 2000), 742 unrelated individuals recruited for behavioral studies (Lueders *et al*, 2002), 621 heavy-smoking men from 206 families recruited for nicotine addiction studies (Feng *et al*, 2004), 516 individuals recruited for behavioral studies (Philibert *et al*, 2009), in addition to treatment-seeking smokers (Bergen *et al*, 2009). In this last study, a combined analysis of 821 treatment-seeking participants of two smoking cessation trials (Lerman *et al*, 2006) did not identify significant associations between rs2072660 and rs2072661 and baseline nicotine dependence as measured by the FTND score (uncorrected  $P$ -values of 0.280 and 0.195 for rs2072660 and rs2072661, respectively) (Bergen *et al*, 2009). Gene-gene interaction analyses did not identify significant interactions of *CHRN2* SNPs (including rs2072660 and rs2072661) with SNPs from *CHRNA4*, *NTRK2*, and *BDNF* in association with smoking status in a sample of 191 unrelated smokers and 191 unrelated nonsmokers (Lou *et al*, 2007), but did identify a significant interaction of *CHRN2* SNPs (including rs2072660 and rs2072661) and *CHRNA4* SNPs with smoking status in a sample of 275 unrelated smokers with FTND scores  $\geq 4$  and 348 unrelated nonsmokers (Li *et al*, 2008).

A large candidate gene association study in 1929 ever smokers that included 119 SNPs at 16 nAChR subunit genes reported significant associations with SNPs at the *CHRN3/CHRNA6*, *CHRNA5/CHRNA3/CHRN4*, and *CHRN4/CHRN4* multigene loci and a dichotomized measure of nicotine dependence (Saccone *et al*, 2007). A follow-up study using the same sample and phenotype, and analyzing a total of 226 SNPs at 16 nAChR subunit genes, reported significant associations with SNPs within the same multigene loci, and nominal association (not associated after multiple test correction) with SNPs at *CHRN1* and at *CHRNA4* (Saccone *et al*, 2009a). We identified nominal associations between *CHRN1* SNPs and the FTND score in our analyses of 430 treatment-seeking smokers, including at some *CHRN1* SNPs studied by Saccone *et al* (2009a) and Lou *et al* (2006); however, these nominal associations became nonsignificant after correction for correlated tests at *CHRN1*.

We identified a significant negative association between rare *CHRNA4* SNPs and the FTND score in the CAST test, which remained significant after permutation testing. Previous results of candidate gene association with *CHRNA4* (Breitling *et al*, 2009b; Feng *et al*, 2004; Hutchison *et al*, 2007; Li *et al*, 2005; Saccone *et al*, 2009a) have suggested that multiple common *CHRNA4* SNPs may be associated with nicotine dependence-related phenotypes. The lack of a significant association of common *CHRNA4* SNPs in this sample with FTND could be because of a lack of power and/or differences in the distribution of common and/or rare SNPs in our sample of treatment-seeking smokers relative to previous samples. Differences in results between different tests of association of rare *CHRNA4* variation with the FTND score in this sample (CAST *vs* WSS) may be

because of the large number of rare variants available at *CHRNA4*, which improves the power of the CAST test. However, the nominal significance observed in the CAST test and the lack of association observed in the WSS and MDMR tests suggest that this result should be regarded as preliminary and in need of replication.

One or more differences in study design, ascertainment criteria, assessment of nicotine dependence, case and control definitions, MAF thresholds, demographic variables, or sample sizes could be contributing to differences in results of association of nAChR subunit gene SNPs with measures of nicotine dependence or smoking status among the various studies. For example, the treatment-seeking smokers in our sample have a minimum past year history of  $\geq 10$  cigarettes smoked per day, are mostly female, have an average age of 49 years and are screened for various physical and mental disorders typical of exclusion criteria for smoking cessation clinical trials. These smoking behavior characteristics differ somewhat from population- or family-based samples ascertained with smoking behavior criteria with lower or higher smoking intensities than those observed among treatment-seeking smokers. It is of interest that FTND scores seemed to be normally distributed (data not shown) in this treatment-seeking smoker sample, as in a previous investigation in this patient population (Swan *et al*, 2003) and in other treatment-seeking smoker samples recently investigated (Bergen *et al*, 2009). We note that the sequencing coverage varied across nAChR subunit genes in this study and that some genes had more common and/or rare SNPs identified than other genes in these analyses. Chance is another factor that could be responsible for association differences among studies.

There were limitations to our analysis. One limitation is the use of saliva-extracted DNA obtained from participants of a smoking cessation trial, in which genotyping performance using DNA extracted from saliva may be reduced by increased nonhuman DNA content (Herraez and Stoneking, 2008). Three genes (namely *CHRNA4*, *CHRNA5*, and *CHRNA7*) exhibited reduced 454 resequencing coverage due to targeted regions of these genes with GC content that exceeds 80% (*CHRNA4*, 5'UTR, exon 1, exon 5, exon 6, 3' UTR; *CHRNA5*, 5'UTR, exon 1, exon 2, exon 5, exon 6, 3' UTR; *CHRNA7*, exon 1). For two of these genes (*CHRNA4* and *CHRNA5*), we supplemented next-generation sequencing with Sanger sequencing protocols, which increased the number of rare and common SNPs available for analysis at these genes. The *CHRNA7* region contains a partial duplication (Gault *et al*, 1998), which may have contributed to reduced resequencing coverage of this gene.

Power analyses to discover *de novo* rare variant associations from sequencing efforts have cited sample sizes from 100 to 10 000 depending on a number of factors, including subject sampling criteria, effect sizes, variant frequency, strength of purifying selection, and the probability of detecting variants (Kryukov *et al*, 2009; Li and Leal, 2008). Including this study, four groups have resequenced 3–10 nAChR genes using 80–675 individuals identifying 55–262 SNPs (Rana *et al*, 2009; Sabatelli *et al*, 2009; Weiss *et al*, 2008). Each of the four studies identified previously unreported variants; however, the power to identify associations with rare variants is limited by the lower sample sizes suggesting the need for larger sample sizes and more

efficient approaches toward identifying rare, potentially functional variants. Small sample sizes leading to underpowered candidate gene studies have been an often-cited reason for inconsistent results across studies (Ioannidis, 2008). In our sample of 430 individuals with a mean (SD) FTND score of 5.2 (2.1), the power of single SNP tests (double sided with  $\alpha = 0.05$ ) to detect  $\beta$ -values of 0.5, 1.0, 1.5, and 2.0 with a MAF = 0.01 and a dominant model is 11, 28, 54, and 78%, whereas the *post hoc* power of the linear regression approach to detect the effect of the common *CHRNA5* variants significantly associated with FTND score is 70%.

In conclusion, there are numerous previous GWASs and candidate gene studies supporting the hypothesis that common variation in several nAChR subunit genes contributes to variation in nicotine dependence as measured by the FTND or by measures of smoking intensity. In our resequencing study of DNA from a sample of treatment-seeking smokers, we found a significant evidence for the contribution of common variants at *CHRNA5* and rare variants at *CHRNA4* to variation in a measure of nicotine dependence. We note that in mouse knockout studies, the  $\beta_2$  and  $\alpha_4$  nicotinic receptor subunit genes have been shown to be necessary and sufficient to observe the reinforcing properties of nicotine (Mineur and Picciotto, 2008), and that the human  $\alpha_4\beta_2$  nAChR is upregulated upon nicotine administration (Buisson and Bertrand, 2001), providing support for the involvement of these two nAChR subunits in the etiology of nicotine dependence. The heteropentameric structure of the nAChR suggests that future gene-gene interaction analyses of nAChR subunit variants will be a useful approach to identify nAChR sequence variation influencing smoking behavior. Analysis of both common and rare variants in the COMPASS samples suggests that the combination of common and rare variants at *CHRNA5*, *CHRNA4*, and at the chr15q25.1 nAChR locus, also contribute to nicotine dependence. The chr15q25.1 and *CHRNA4* findings are concordant with multiple previous association results in both GWASs and candidate gene studies, whereas the *CHRNA5* finding may reflect the common SNP association findings already observed in this sample with SNP-wise analysis. Resequencing of DNA from nAChR subunit genes targeting coding and regulatory regions, using larger samples sizes and multiple approaches to statistical genetic analysis, including analyses of interaction and synthetic association hypotheses, will be necessary to further characterize the relationship between common and rare nAChR subunit gene sequence variation and smoking-related behaviors in treatment-seeking participants of smoking cessation trials.

## ACKNOWLEDGEMENTS

This research was supported by R01 CA071358, U01 DA020830, and a Material Transfer and Research Agreement between SRI International and Perlegen Sciences. We thank the participants of the COMPASS smoking cessation clinical trial (CA071358, PI Gary Swan). We thank David V Conti for helpful comments, and we acknowledge the efforts of Ondrej Libiger at Scripps Genomic Medicine, Susan Zbikowski and Tim McAfee of Free & Clear, Julie Richards



and the staff of the Group Health Research Institute's Survey Research Program, and Martha Agreda, Mary McElroy, Gaye Courtney, Peggy Giacalone, and Trish McLeod of SRI International. We thank the three anonymous reviewers for their thoughtful and detailed comments and suggestions.

## DISCLOSURE

*Related to this work.* The COMPASS study was funded by the National Cancer Institute grant R01 CA071358 and is registered at Clinicaltrials.gov (NCT00301145). Varenicline and nominal support for recruiting COMPASS participants was provided by Pfizer. Neither entity had any role in the COMPASS study design or in the collection, analysis, and interpretation of COMPASS data. Authors Wessel, McDonald, Hinds, Stokowski, Javitz, Kennemer, Krasnow, Dirks, Hardin, Pitts, Michel, Jack, Ballinger, McClure, Swan, and Bergen declare that all financial and material support for this work was provided by their primary employer. *Related to all other compensation.* Authors Wessel, McDonald, Stokowski, Javitz, Krasnow, Dirks, Hardin, Michel, Jack, and McClure declare that, except for income provided from their primary employer, no financial support or compensation has been received from any individual or corporate entity over the past 3 years for research or professional services and there are no personal financial holdings that could be perceived as constituting a potential conflict of interest. Dr Hinds is currently employed by 23 and Me, Dr Kennemer is currently employed by Complete Genetics as Genomics Applications Analyst, Dr Pitts is currently employed by Pfizer, in its Applied Quantitative Genotherapeutics group, and Dr Ballinger is currently employed by Complete Genomics as Vice President of Genomics. Dr Swan received financial support from Pfizer to attend a 1-day advisory meeting in 2008. In the past 3 years, Dr Bergen has received compensation for professional services from the National Institutes of Health and the National Coalition for Health Professional Education in Genetics.

## REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* 215: 403–410.
- Baker TB, Weiss RB, Bolt D, von Niederhausern A, Fiore MC, Dunn DM *et al* (2009). Human neuronal acetylcholine receptor A5-A3-B4 haplotypes are associated with multiple nicotine dependence phenotypes. *Nicotine Tob Res* 11: 785–796.
- Barrett JC, Fry B, Maller J, Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–265.
- Bergen AW, Caporaso N (1999). Cigarette smoking. *J Natl Cancer Inst* 91: 1365–1375.
- Bergen AW, Conti DV, Van Den Berg D, Lee W, Liu J, Li D *et al* (2009). Dopamine genes and nicotine dependence in treatment-seeking and community smokers. *Neuropsychopharmacology* 34: 2252–2264.
- Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Grucza RA, Xuei X *et al* (2008). Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry* 165: 1163–1171.
- Bodmer W, Bonilla C (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40: 695–701.
- Breitling LP, Dahmen N, Mittelstrass K, Illig T, Rujescu D, Raum E *et al* (2009a). Smoking cessation and variations in nicotinic acetylcholine receptor subunits alpha-5, alpha-3, and beta-4 genes. *Biol Psychiatry* 65: 691–695.
- Breitling LP, Dahmen N, Mittelstrass K, Rujescu D, Gallinat J, Fehr C *et al* (2009b). Association of nicotinic acetylcholine receptor subunit alpha 4 polymorphisms with nicotine dependence in 5500 Germans. *Pharmacogenomics J* 9: 219–224.
- Broms U, Madden PA, Heath AC, Pergadia ML, Shiffman S, Kaprio J (2007). The Nicotine Dependence Syndrome Scale in Finnish smokers. *Drug Alcohol Depend* 89: 42–51.
- Buisson B, Bertrand D (2001). Chronic exposure to nicotine upregulates the human (alpha)4((beta)2 nicotinic acetylcholine receptor function. *J Neurosci* 21: 1819–1829.
- Changeux JP (2009). Nicotinic receptors and nicotine addiction. *C R Biol* 332: 421–425.
- Chen GB, Payne TJ, Lou XY, Ma JZ, Zhu J, Li MD (2008). Association of amyloid precursor protein-binding protein, family B, member 1 with nicotine dependence in African and European American smokers. *Hum Genet* 124: 393–398.
- Chen X, Chen J, Williamson VS, An SS, Hetta JM, Aggen SH *et al* (2009). Variants in nicotinic acetylcholine receptors alpha5 and alpha3 increase risks to nicotine dependence. *Am J Med Genet B Neuropsychiatr Genet* 150B: 926–933.
- Conneely KN, Boehnke M (2007). So many correlated tests, so little time! rapid adjustment of P values for multiple correlated tests. *Am J Hum Genet* 81: 1158–1168.
- Conti DV, Lee W, Li D, Liu J, Van Den Berg D, Thomas PD *et al* (2008). Nicotinic acetylcholine receptor beta2 subunit gene implicated in a systems-based candidate gene study of smoking cessation. *Hum Mol Genet* 17: 2834–2848.
- Craig DW, Pearson JW, Szelinger S, Sekar A, Redman M, Corneveaux JJ *et al* (2008). Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods* 5: 887–893.
- Dani JA, Bertrand D (2007). Nicotinic acetylcholine receptors and nicotinic cholinergic mechanisms of the central nervous system. *Annu Rev Pharmacol Toxicol* 47: 699–729.
- Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB (2010). Rare variants create synthetic genome-wide associations. *PLoS Biol* 8: e1000294.
- Drgon T, Johnson C, Walther D, Albino AP, Rose JE, Uhl GR (2009a). Genome-wide association for smoking cessation success: participants in a trial with adjunctive denicotinized cigarettes. *Mol Med* 15: 268–274.
- Drgon T, Montoya I, Johnson C, Liu QR, Walther D, Hamer D *et al* (2009b). Genome-wide association for nicotine dependence and smoking cessation success in NIH research volunteers. *Mol Med* 15: 21–27.
- Ehringer MA, Clegg HV, Collins AC, Corley RP, Crowley T, Hewitt JK *et al* (2007). Association of the neuronal nicotinic receptor beta2 subunit gene (CHRNA2) with subjective responses to alcohol and nicotine. *Am J Med Genet B Neuropsychiatr Genet* 144B: 596–604.
- Etter JF, Hoda JC, Perroud N, Munafo M, Buresi C, Duret C *et al* (2009). Association of genes coding for the alpha-4, alpha-5, beta-2 and beta-3 subunits of nicotinic receptors with cigarette smoking and nicotine dependence. *Addict Behav* 34: 772–775.
- Fagerstrom K, Furberg H (2008). A comparison of the Fagerstrom test for nicotine dependence and smoking prevalence across countries. *Addiction* 103: 841–845.
- Feng Y, Niu T, Xing H, Xu X, Chen C, Peng S *et al* (2004). A common haplotype of the nicotine acetylcholine receptor alpha 4 subunit gene is associated with vulnerability to nicotine addiction in men. *Am J Hum Genet* 75: 112–121.
- Frazer KA, Murray SS, Schork NJ, Topol EJ (2009). Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10: 241–251.

- Freathy RM, Ring SM, Shields B, Galobardes B, Knight B, Weedon MN *et al* (2009). A common genetic variant in the 15q24 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNA4) is associated with a reduced ability of women to quit smoking in pregnancy. *Hum Mol Genet* **18**: 2922–2927.
- Gauderman WJ (2002). Sample size requirements for matched case-control studies of gene-environment interaction. *Stat Med* **21**: 35–50.
- Gault J, Robinson M, Berger R, Drebing C, Logel J, Hopkins J *et al* (1998). Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene (CHRNA7). *Genomics* **52**: 173–185.
- Greenbaum L, Kanyas K, Karni O, Merbl Y, Olender T, Horowitz A *et al* (2006). Why do young women smoke? I. Direct and interactive effects of environment, psychological characteristics and nicotinic cholinergic receptor genes. *Mol Psychiatry* **11**: 312–322, 223.
- Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY *et al* (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* **10**: R32.
- Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO (1991). The Fagerstrom test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict* **86**: 1119–1127.
- Herraez DL, Stoneking M (2008). High fractions of exogenous DNA in human buccal samples reduce the quality of large-scale genotyping. *Anal Biochem* **383**: 329–331.
- Hinds DA, Stokowski RP, Patil N, Konvicka K, Kershenovich D, Cox DR *et al* (2004). Matching strategies for genetic association studies in structured populations. *Am J Hum Genet* **74**: 317–325.
- Hoft NR, Corley RP, McQueen MB, Schlaepfer IR, Huizinga D, Ehringer MA (2009). Genetic association of the CHRNA6 and CHRNA3 genes with tobacco dependence in a nationally representative sample. *Neuropsychopharmacology* **34**: 698–706.
- Hutchison KE, Allen DL, Filbey FM, Jepson C, Lerman C, Benowitz NL *et al* (2007). CHRNA4 and tobacco dependence: from gene regulation to treatment outcome. *Arch Gen Psychiatry* **64**: 1078–1086.
- Ioannidis JP (2008). Why most discovered true associations are inflated. *Epidemiology* **19**: 640–648.
- Keskitalo K, Broms U, Heliovaara M, Ripatti S, Surakka I, Perola M *et al* (2009). Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNA4) on chromosome 15. *Hum Mol Genet* **18**: 4007–4012.
- Kryukov GV, Shpunt A, Stamatoyannopoulos JA, Sunyaev SR (2009). Power of deep, all-exon resequencing for discovery of human trait genes. *Proc Natl Acad Sci USA* **106**: 3871–3876.
- Le Marchand L, Derby KS, Murphy SE, Hecht SS, Hatsukami D, Carmella SG *et al* (2008). Smokers with the CHRNA lung cancer-associated variants are exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine. *Cancer Res* **68**: 9137–9140.
- Lerman C, Jepson C, Wileyto EP, Epstein LH, Rukstalis M, Patterson F *et al* (2006). Role of functional genetic variation in the dopamine D2 receptor (DRD2) in response to bupropion and nicotine replacement therapy for tobacco dependence: results of two randomized clinical trials. *Neuropsychopharmacology* **31**: 231–242.
- Lessov CN, Martin NG, Statham DJ, Todorov AA, Slutske WS, Bucholz KK *et al* (2004). Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychol Med* **34**: 865–879.
- Li B, Leal SM (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* **83**: 311–321.
- Li MD, Beuten J, Ma JZ, Payne TJ, Lou XY, Garcia V *et al* (2005). Ethnic- and gender-specific association of the nicotinic acetylcholine receptor alpha4 subunit gene (CHRNA4) with nicotine dependence. *Hum Mol Genet* **14**: 1211–1219.
- Li MD, Cheng R, Ma JZ, Swan GE (2003). A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction* **98**: 23–31.
- Li MD, Lou XY, Chen G, Ma JZ, Elston RC (2008). Gene-gene interactions among CHRNA4, CHRNA2, BDNF, and NTRK2 in nicotine dependence. *Biol Psychiatry* **64**: 951–957.
- Li MD, Xu Q, Lou XY, Payne TJ, Niu T, Ma JZ (2009). Association and interaction analysis of variants in CHRNA5/CHRNA3/CHRNA4 gene cluster with nicotine dependence in African and European Americans. *Am J Med Genet B Neuropsychiatr Genet* **153B**: 745–756.
- Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L *et al* (2010). Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* **42**: 436–440.
- Liu YZ, Pei YF, Guo YF, Wang L, Liu XG, Yan H *et al* (2009). Genome-wide association analyses suggested a novel mechanism for smoking behavior regulated by IL15. *Mol Psychiatry* **14**: 668–680.
- Lou XY, Chen GB, Yan L, Ma JZ, Mangold JE, Zhu J *et al* (2008). A combinatorial approach to detecting gene-gene and gene-environment interactions in family studies. *Am J Hum Genet* **83**: 457–467.
- Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC *et al* (2007). A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *Am J Hum Genet* **80**: 1125–1137.
- Lou XY, Ma JZ, Payne TJ, Beuten J, Crew KM, Li MD (2006). Gene-based analysis suggests association of the nicotinic acetylcholine receptor beta1 subunit (CHRNA1) and M1 muscarinic acetylcholine receptor (CHRM1) with vulnerability for nicotine dependence. *Hum Genet* **120**: 381–389.
- Lueders KK, Hu S, McHugh L, Myakishev MV, Sirota LA, Hamer DH (2002). Genetic and functional analysis of single nucleotide polymorphisms in the beta2-neuronal nicotinic acetylcholine receptor gene (CHRNA2). *Nicotine Tob Res* **4**: 115–125.
- Madsen BE, Browning SR (2009). A group-wise association test for rare mutations using a weighted sum statistic. *PLoS Genet* **5**: e1000384.
- Maes HH, Sullivan PF, Bulik CM, Neale MC, Prescott CA, Eaves LJ *et al* (2004). A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychol Med* **34**: 1251–1261.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.
- Mineur YS, Picciotto MR (2008). Genetics of nicotinic acetylcholine receptors: relevance to nicotine addiction. *Biochem Pharmacol* **75**: 323–333.
- Nievergelt CM, Libiger O, Schork NJ (2007). Generalized analysis of molecular variance. *PLoS Genet* **3**: e51.
- Nishita DM, Jack LM, McElroy M, McClure JB, Richards J, Swan GE *et al* (2009). Clinical trial participant characteristics and saliva and DNA metrics. *BMC Med Res Methodol* **9**: 71.
- Perkins KA, Lerman C, Coddington S, Jetton C, Karelitz JL, Wilson A *et al* (2008a). Gene and gene by sex associations with initial sensitivity to nicotine in nonsmokers. *Behav Pharmacol* **19**: 630–640.
- Perkins KA, Lerman C, Grotenthaler A, Ciccocioppo MM, Milanak M, Conklin CA *et al* (2008b). Dopamine and opioid gene variants are associated with increased smoking reward and reinforcement owing to negative mood. *Behav Pharmacol* **19**: 641–649.
- Perkins KA, Lerman C, Mercincavage M, Fonte CA, Briski JL (2009). Nicotinic acetylcholine receptor beta2 subunit (CHRNA2) gene and short-term ability to quit smoking in response to nicotine patch. *Cancer Epidemiol Biomarkers Prev* **18**: 2608–2612.

- Philibert RA, Todorov A, Andersen A, Hollenbeck N, Gunter T, Heath A *et al* (2009). Examination of the nicotine dependence (NICSNP) consortium findings in the Iowa adoption studies population. *Nicotine Tob Res* **11**: 286–292.
- Piper ME, McCarthy DE, Baker TB (2006). Assessing tobacco dependence: a guide to measure evaluation and selection. *Nicotine Tob Res* **8**: 339–351.
- Rana BK, Wessel J, Mahboubi V, Rao F, Haeller J, Gayen JR *et al* (2009). Natural variation within the neuronal nicotinic acetylcholine receptor cluster on human chromosome 15q24: influence on heritable autonomic traits in twin pairs. *J Pharmacol Exp Ther* **331**: 419–428.
- Rigbi A, Kanyas K, Yakir A, Greenbaum L, Pollak Y, Ben-Asher E *et al* (2008). Why do young women smoke? V. Role of direct and interactive effects of nicotinic cholinergic receptor gene variation on neurocognitive function. *Genes Brain Behav* **7**: 164–172.
- Sabatelli M, Eusebi F, Al-Chalabi A, Conte A, Madia F, Luigetti M *et al* (2009). Rare missense variants of neuronal nicotinic acetylcholine receptor altering receptor function are associated with sporadic amyotrophic lateral sclerosis. *Hum Mol Genet* **18**: 3997–4006.
- Saccone NL, Saccone SF, Hinrichs AL, Stitzel JA, Duan W, Pergadia ML *et al* (2009a). Multiple distinct risk loci for nicotine dependence identified by dense coverage of the complete family of nicotinic receptor subunit (CHRN) genes. *Am J Med Genet B Neuropsychiatr Genet* **150B**: 453–466.
- Saccone NL, Wang JC, Breslau N, Johnson EO, Hatsukami D, Saccone SF *et al* (2009b). The CHRNA5-CHRNA3-CHRNA4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res* **69**: 6848–6856.
- Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA *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.
- Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463–5467.
- Schork NJ, Murray SS, Frazer KA, Topol EJ (2009). Common vs rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* **19**: 212–219.
- Schork NJ, Wessel J, Malo N (2008). DNA sequence-based phenotypic association analysis. *Adv Genet* **60**: 195–217.
- Sherva R, Wilhelmsen K, Pomerleau CS, Chasse SA, Rice JP, Snedecor SM *et al* (2008). Association of a single nucleotide polymorphism in neuronal acetylcholine receptor subunit alpha 5 (CHRNA5) with smoking status and with 'pleasurable buzz' during early experimentation with smoking. *Addiction* **103**: 1544–1552.
- Silverman MA, Neale MC, Sullivan PF, Harris-Kerr C, Wormley B, Sadek H *et al* (2000). Haplotypes of four novel single nucleotide polymorphisms in the nicotinic acetylcholine receptor beta2-subunit (CHRNA2) gene show no association with smoking initiation or nicotine dependence. *Am J Med Genet* **96**: 646–653.
- Spitz MR, Amos CI, Dong Q, Lin J, Wu X (2008). The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. *J Natl Cancer Inst* **100**: 1552–1556.
- Stevens VL, Bierut LJ, Talbot JT, Wang JC, Sun J, Hinrichs AL *et al* (2008). Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol Biomarkers Prev* **17**: 3517–3525.
- Stolerman IP, Jarvis MJ (1995). The scientific case that nicotine is addictive. *Psychopharmacology (Berl)* **117**: 2–10; discussion 14–20.
- Swan GE, McAfee T, Curry SJ, Jack LM, Javitz H, Dacey S *et al* (2003). Effectiveness of bupropion sustained release for smoking cessation in a health care setting: a randomized trial. *Arch Intern Med* **163**: 2337–2344.
- Swan GE, McClure JB, Jack LM, Zbikowski SM, Javitz H, Catz SL *et al* (2010). Behavioral counseling and varenicline treatment for smoking cessation. *Am J Prev Med* **38**: 482–490.
- TAG (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**: 441–447.
- Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP *et al* (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**: 638–642.
- Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F *et al* (2010). Sequence variants at CHRNA3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet* **42**: 448–453.
- Vink JM, Willemsen G, Boomsma DI (2005). Heritability of smoking initiation and nicotine dependence. *Behav Genet* **35**: 397–406.
- Wang J, Li MD (2010). Common and unique biological pathways associated with smoking initiation/progression, nicotine dependence, and smoking cessation. *Neuropsychopharmacology* **35**: 702–719.
- Wang JC, Gruzca R, Cruchaga C, Hinrichs AL, Bertelsen S, Budde JP *et al* (2009). Genetic variation in the CHRNA5 gene affects mRNA levels and is associated with risk for alcohol dependence. *Mol Psychiatry* **14**: 501–510.
- Wang K, Dickson SP, Stolle CA, Krantz ID, Goldstein DB, Hakonarson H (2010). Interpretation of association signals and identification of causal variants from genome-wide association studies. *Am J Hum Genet* **86**: 730–742.
- Weiss RB, Baker TB, Cannon DS, von Niederhausern A, Dunn DM, Matsunami N *et al* (2008). A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genet* **4**: e1000125.
- Wessel J, Schork NJ (2006). Generalized genomic distance-based regression methodology for multilocus association analysis. *Am J Hum Genet* **79**: 792–806.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (<http://www.nature.com/npp>)