Use of polygenic risk scores of nicotine metabolism in predicting smoking behaviors

Li-Shiun Chen^{1,2}, Sarah M Hartz¹, Timothy B Baker³, Yinjiao Ma¹, Nancy L Saccone⁴ & Laura J Bierut^{*,1,2}

¹Department of Psychiatry, Washington University School of Medicine, St Louis, MO 63110, USA

²Siteman Cancer Center, Washington University in St Louis, St Louis, MO 63110, USA

³Department of Medicine, University of Wisconsin School of Medicine & Public Health, Madison, WI 53711, USA

⁴Department of Genetics, Washington University School of Medicine, St Louis, MO 63110, USA

*Author for correspondence: Tel.: +1 314 362 3492; laura@wustl.edu

Aim: This study tests whether polygenic risk scores (PRSs) for nicotine metabolism predict smoking behaviors in independent data. **Materials & methods:** Linear regression, logistic regression and survival analyses were used to analyze nicotine metabolism PRSs and nicotine metabolism, smoking quantity and smoking cessation. **Results:** Nicotine metabolism PRSs based on two genome wide association studies (GWAS) meta-analyses significantly predicted nicotine metabolism biomarkers (R^2 range: 9.2–16%; minimum $p = 7.6 \times 10^{-8}$). The GWAS top hit variant rs56113850 significantly predicted nicotine metabolism biomarkers (R^2 range: 14–17%; minimum $p = 4.4 \times 10^{-8}$). There was insufficient evidence for these PRSs predicting smoking quantity and smoking cessation. **Conclusion:** Results suggest that nicotine metabolism PRSs based on GWAS meta-analyses predict an individual's nicotine metabolism, so does use of the top hit variant. We anticipate that PRSs will enter clinical medicine, but additional research is needed to develop a more comprehensive genetic score to predict smoking behaviors.

First draft submitted: 21 May 2018; Accepted for publication: 10 October 2018; Published online: 16 November 2018

Keywords: *CYP2A6* • nicotine metabolism • polygenic risk scores • smoking cessation

Variation in nicotine metabolism influences smoking behaviors. Different levels of nicotine metabolism, and genetic variation that encodes the primary nicotine metabolism enzyme, cytochrome P450 2A6 (*CYP2A6*), have been shown as associated with smoking behaviors, especially cigarette consumption as well as smoking cessation [1–4]. The gene *CYP2A6* is highly polymorphic, with variation that reduces the enzymatic function leading to significantly slower rates of nicotine metabolism. Multiple common variants define levels of nicotine metabolism in European populations [5]. In fact, the majority of interindividual variations in the metabolism of nicotine to its first major metabolite, cotinine, can be explained by seven genetic polymorphisms among European-Americans [6].

With regards to the prediction of smoking cessation, many of the studies linking nicotine metabolism and cessation treatment have examined a proxy for *CYP2A6* activity, nicotine metabolite ratio (NMR), the ratio of two stable nicotine metabolites: 3-hydroxycotinine to cotinine, measured in the blood of current smokers [7–11]. Several studies have reported meaningful association between NMR and cessation [7–9], and there is evidence of an interaction between NMR and type of cessation treatment. Specifically, slow nicotine metabolizers (low NMRs) are more likely to become abstinent than fast metabolizers when both receive either placebo treatment or nicotine replacement therapy (NRT) [7–9]. Conversely, varenicline was more effective than NRT among fast metabolizers [12]. More recent evidence suggests that the effect of NMR on smoking cessation success is complex [13].

An alternative method to characterize metabolism is to use genetic variation in *CYP2A6*, which in turn predicts variation in nicotine metabolism [6,14]. In an experiment using labeled nicotine, we determined conversion of nicotine to cotinine as defined as the proportion, cotinine/[nicotine + cotinine], to develop a predictive model of nicotine metabolism based on *CYP2A6* genotypes, referred to as the *CYP2A6* metric [6,14]. *CYP2A6* variation included in this model explained 70% ($R^2 = 0.7$) of the variance in the rapid conversion of nicotine to cotinine among European-Americans. Metabolism estimates predicted by the model were significantly correlated with

self-reported cigarettes per day (CPD) [1,6] and exhaled carbon monoxide (CO) [15]. The *CYP2A6* metric has been shown to predict smoking quantity, smoking cessation and response to pharmacotherapy such as NRT. In particular, we reported that the effect of NRT varies with metabolism speed as defined by this *CYP2A6* metric: fast metabolizers were more likely to relapse in a smoking cessation trial in the placebo arm and fast, but not slow, metabolizers benefited from nicotine replacement therapy [16].

Two recent genome wide association studies (GWAS) meta-analyses of nicotine metabolism measured by the NMR identified multiple associated genetic markers. Loukola *et al.* [17] examined the phenotype of NMR and identified 719 genome-wide significant findings that were all on chromosome 19, the chromosome where *CYP2A6* lies. Similarly, Baurley *et al.* [18] examined the phenotype of NMR and identified 11 genome-wide significant findings, which were also all on chromosome 19. With these recent GWAS for NMR, we now have the opportunity to use polygenic risk scores (PRSs) to capture the effect of variation in *CYP2A6* more comprehensively. PRSs are created from the multiple genetic variants identified in these NMR GWASs. Genetic variants are given an associated weight based on the discovery GWAS and these weights then provide a predicted metabolic rate in other samples. The composite PRSs, which incorporate a large number of variants, have higher predictive performance than an individual genetic variant [19].

Nicotine metabolism is complex and the metabolite biomarkers synthesize this complexity into a comprehensive score. In addition, genotyping for the *CYP2A6* region is difficult because of the genetic architecture of this region which is easily captured in GWAS arrays. For example, the *CYP2A6* metric requires assessment of copy number variation (CNV) not routinely included in standard GWAS arrays. However, many studies have standard genotype information for their subjects, but have not conducted specific assays of NMR. It has also been reported that *CYP2A6** alleles can be used to predict variance of experimental NMR [20]. Given the new GWAS evidence of the biomarker NMR, we can determine the predictive utility of single variants or polygenic scores based on standard GWAS genotyping/imputation in estimating nicotine metabolism. How well the PRS for NMR can predict an individual's nicotine metabolism and thus smoking behaviors could encourage research in many genetics studies without specific nicotine metabolism biomarker data and potentially inform clinical care in the future.

Our goal in this study is to examine the different nicotine metabolism PRSs that capture *CYP2A6* variation to test whether these predict nicotine metabolism in an independent sample as well as smoking behaviors such as smoking quantity and smoking cessation in different populations. Therefore, using data from a multi-armed cessation trial and three community-based, cross-sectional studies, we tested the following hypotheses: do nicotine metabolism PRSs predict nicotine metabolism biomarkers; do nicotine metabolism PRSs predict smoking quantity; and do nicotine metabolism PRSs predict smoking cessation, defined in the treatment trial as smoking abstinence at end of treatment and latency to smoking relapse, and defined in community-based, cross-sectional studies as age of smoking cessation. This research was designed to reveal the utility of PRSs based on GWAS meta-analyses of nicotine metabolism in predicting smoking behaviors in different study populations. The results have clear relevance in the development of a genetically informed, personalized approach to prevention and treatment of tobacco use disorder.

Materials & methods

Datasets

Four datasets were used for these analyses (Supplementary Table 1). We restricted analyses to individuals of European descent (N = 9199), consistent with the population used in the GWAS for nicotine metabolism from which the PRSs are derived. We confirmed ancestry through principal component analyses. The Collaborative Genetic Study of Nicotine Dependence (COGEND) and Genetic and Environmental Risks for Smoking Characteristics and Cessation (GERS) [21,22] were ascertained for tobacco use disorder (n = 1866). The Atherosclerosis Risk in Communities (ARIC; n = 5063) and Multi-Ethnic Study of Atherosclerosis (MESA; n = 1203) samples were ascertained for atherosclerosis. The University of Wisconsin Transdisciplinary Tobacco Use Research Center (UW-TTURC) sample is from a randomized, placebo-controlled smoking cessation trial (n = 1067) [23]. Each study obtained informed consent from all participants and was approved from the appropriate institutional review boards.

Recruitment

Collaborative genetic study of nicotine dependence & genetic & environmental risks for smoking characteristics & cessation

COGEND was designed to discover genes that alter risk for nicotine dependence. Community-based recruitment enrolled participants with nicotine dependence and control subjects who smoked without nicotine dependence in St Louis, MO, USA, and Detroit, MI, USA in 2002–2007. All participants were 25–44 years old and spoke English. Subjects with nicotine dependence were defined as current smokers with a Fagerstrom Test for Nicotine Dependence (FTND) score ≥ 4 . Control status was defined as ever smoking ≥ 100 cigarettes, but with a lifetime FTND score ≤ 1 . A subgroup from the COGEND study measured nicotine metabolism biomarkers including nicotine, cotinine and 3-hydroxycotinine ($n = 184$) [6]. Another subgroup of tobacco use disorder cases from the COGEND study ($n = 479$) was followed-up via telephone approximately 8 years later to assess smoking cessation outcomes as part of the Genetic and GERS study.

Atherosclerosis risk in communities

The ARIC Study, sponsored by the National Heart, Lung and Blood Institute (NHLBI), is a longitudinal epidemiologic study to investigate the risk factors, medical care, natural history, and etiology of atherosclerosis and cardiovascular disease. Participants aged 45–64, at baseline, with and without atherosclerosis and cardiovascular disease were enrolled in Forsyth County, NC; Jackson, MS; the northwest suburbs of Minneapolis, MN; and Washington County, MD (USA) between 1987 and 1989 and followed-up yearly by telephone. A substudy, part of the Gene Environment Association Studies initiative (GENEVA, <http://www.genevastudy.org>), funded by the trans-NIH Genes, Environment and Health Initiative (GEI), contains genotype and phenotype data. The aim of this study is to identify and characterize factors that contribute to atherosclerosis and cardiovascular disease.

Multi-ethnic study of atherosclerosis

The MESA is a longitudinal family study, sponsored by the National Heart, Lung and Blood Institute (NHLBI) and by the National Center for Research Resources, to identify and characterize genes and risk factors contributing to cardiovascular disease and the cardiovascular disease progression. Diverse, asymptomatic men and women, aged 45–84 were recruited from six field centers across the USA. Each participant received an extensive physical exam and examination of risk factors such as standard coronary risk factors, sociodemographic factors, lifestyle factors and psychosocial factors. Participants were followed to study cardiovascular disease events and cardiovascular disease event progression, including clinical morbidity and mortality.

University of Wisconsin Transdisciplinary Tobacco Use Research Center

Study participants were from a UW-TTURC randomized, placebo-controlled smoking cessation trial [23]; they were ≥ 18 years of age, smoked ≥ 10 cigarettes per day (CPD), and were motivated to quit smoking. Participants completed baseline assessments of demographics, smoking history and nicotine dependence including the FTND [24], prior to randomization. To verify the smoking status and to estimate the level of dependence, participants were asked to give a breath sample for alveolar CO assay. The treatment phase in the trial lasted 8 weeks. Participants (European–American: $N = 1067$) were randomly assigned to: placebo ($n = 130$), bupropion SR ($n = 186$), nicotine replacement therapy ($n = 558$), or combined bupropion and nicotine replacement therapy ($n = 193$). The pharmacotherapies were: placebo medication; bupropion SR (150 mg twice daily for 9 weeks total: 1-week prequit plus 8 weeks postquit); nicotine replacement therapy including nicotine lozenge (2 or 4 mg for 12-week postquit); nicotine patch (24-h patch; 21, 14, and 7 mg; titrated down during 8-week postquit); the combination of nicotine patch and nicotine lozenge (dosed as listed above); or the combination of bupropion SR and nicotine lozenge (dosed as listed above). All participants received six brief (10 min) individual counseling sessions.

Assessments

Nicotine metabolism biomarkers

In the COGEND subsample, we assessed plasma concentrations of nicotine and its metabolites and also deuterated (D2) nicotine metabolites after administration of D2-nicotine in both current ($n = 82$) and former smokers ($n = 102$). For analysis, we defined three metabolism markers: the NMR for current smokers; experimental NMR (eNMR) as D2-3-hydroxycotinine/D2-cotinine for both current and former smokers at 4 hours after nicotine administration; and experimentally derived conversion of nicotine to cotinine as defined by the metabolism

proportion, D2-cotinine/(D2-nicotine + D2-cotinine), at 30 min after nicotine administration [6]. Nicotine metabolism biomarkers are log-transformed as dependent variables.

Smoking quantity

Smoking quantity when subjects smoked regularly was assessed with self-reported cigarettes per day (CPD), defined as a 4-level ordered trait ($CPD \leq 10$; $11 \leq CPD \leq 20$; $21 \leq CPD \leq 30$ and $CPD \geq 31$, coded as 0, 1, 2 and 3, respectively).

Smoking cessation

For the community-based, cross-sectional studies (ARIC, MESA), the primary smoking cessation phenotype was time (in years) from birth to age of smoking cessation defined as self-reported age of stopping smoking among ever smokers. Because most participants initiated smoking within a narrow age range in adolescence [25], age of smoking cessation captures the ranking of smoking duration among participants.

For the smoking cessation treatment trial (UW-TTUTC), biochemically confirmed 7-day point prevalence abstinence was assessed at end-of-treatment (8-week postquit). All of participants' self-reports of abstinence during study visits were confirmed by an expired CO level of less than 10 parts per million. Relapse was defined as any smoking on seven consecutive days after the target quit date. Time of relapse was determined via timeline follow-up assessment [26,27] and was available for 1015 subjects. For the follow-up study of community smokers (GERS), relapse was defined as any smoking on seven consecutive days after the self-reported quit date.

Genotyping

COGEND & GERS

COGEND samples were genotyped on either the Illumina Human1M (dbGaP accession number phs000092.v1.p1; Illumina, CA, USA) or the Illumina 2.5 M (as part of dbGaP accession number phs000404.v1.p1) platforms. These datasets were combined and genotype data from the intersection of the 1 M and 2.5 M platforms were used as the basis for imputation [28].

ARIC

Genotyping was performed at the Broad Institute of MIT and Harvard, a GENEVA genotyping center. Data cleaning and harmonization were performed at the GEI-funded GENEVA Coordinating Center at the University of Washington (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000090.v1.p1).

MESA

DNA was extracted and lymphocytes cryopreserved to allow genetic studies. Genotyping was performed by Affymetrix (CA, USA) and the Broad Institute of Harvard and MIT (MA, USA) using the Affymetrix Genome-Wide Human SNPArray 6.0.

UW-TTUTC

Genotyping of the UW-TTUTC sample was performed by the Center for Inherited Disease Research at Johns Hopkins University using the Illumina Omni2.5 microarray (www.illumina.com). Genotype data were cleaned by the GENEVA Coordinating Center at the University of Washington.

Imputation

All four studies were imputed using the University of Michigan Imputation server (<https://imputationserver.sph.umich.edu>) with 1000 Genomes Phase III version 5 [29] as the reference panel. Prior to uploading genotypes for imputation, maps were updated to Genome Reference Consortium Build 37 (GrCh37) and strand alignment was confirmed.

Statistical analyses

We used standard methods to develop the PRSs [30,31]. The GWAS-level significant variants ($p < 5 \times 10^{-8}$) on chromosome 19 from two NMR Genome Wide Association Study meta-analyses [17,18] were used to generate NMR PRSs for participants in the four independent datasets using the genotype weights from the GWAS results. There was no overlap of participants between these four datasets and the Loukola *et al.* and Baurley *et al.* studies [17,18].

We applied the summary statistics from the NMR meta-analyses European samples to variants with imputation info ≥ 0.9 and allele frequency ≥ 0.02 . Hard call genotypes were then constructed for the PRS analyses. NMR PRSs were calculated for p-value threshold (5×10^{-8}) in NMR meta-analyses using a modified version of PRSice [32], an R language wrapper script using second-generation PLINK [33,34]. SNPs in each dataset were pruned by PRSice (version 1.23) [32] using p-value-informed linkage disequilibrium clumping; $R^2 < 0.10$ in a 500-kb window, collapsed to the most significant variant [30,31]. The PRSs are weighted scores of GWAS level hits after LD pruning and minor allele frequency filter of 0.02. A total of 27 SNPs were used in PRS from Loukola *et al.* [17], and 10 SNPs were used in PRS from Baurley *et al.* [18]. The detailed SNP information and weights are included in Supplementary Tables 2 and 3. PRS quartiles were used as predictors for our primary analyses.

Our first standardized analyses involved testing and evaluating within each dataset the associations between the NMR PRS and each of the smoking behavior phenotypes: nicotine metabolism biomarkers, smoking quantity and smoking cessation.

In each dataset and for each p-value threshold used for the NMR meta-analyses results, the NMR PRSs were regressed against the smoking behavior phenotypes using linear regression, logistic regression or Cox regression. Age, sex and the first two population stratification principal components were included as covariates to adjust for their potential effects on the outcomes. We have adjusted the p-value for statistical significance to 0.0083 based on our testing of two PRSs in three outcomes (nicotine metabolism, smoking quantity, and smoking cessation).

The proportion of variance explained (R^2) by each NMR PRS was computed by comparing the regression model with age, sex, and two principal components to the regression model that includes the NMR PRS variable in addition to age, sex, and principal components. In addition, the proportion of variance explained (R^2) by the top hit (rs56113850) in both Loukola *et al.* [17] and Baurley *et al.* [18] papers was computed.

Results

We studied smoking behaviors in three community-based, cross-sectional studies (COGEND, ARIC, MESA) and one smoking cessation trial (UW-TTUTC). This design allowed us to examine smoking behaviors in general population smokers in the community as well as smokers motivated to quit in a treatment trial. Demographic descriptions for all studies are given in Supplementary Table 1.

Polygenic risk scores based on nicotine metabolism GWAS predict nicotine metabolism biomarkers

In a subsample of COGEND, three nicotine metabolism biomarkers were available: NMR, NMR of experimentally ingested nicotine (eNMR), assessed at 4 hour after experimental ingestion of nicotine, and nicotine metabolism proportion (cotinine/[nicotine + cotinine]) assessed 30 min after experimentally ingested nicotine. These three nicotine metabolism biomarkers are moderately correlated at 0.42–0.79 range; $p < 0.0001$ (Supplementary Table 4).

We examined the correlation among the NMR PRS based on the Loukola *et al.* GWAS, the NMR PRS based on the Baurley *et al.* GWAS, and the CYP2A6 activity metric based on haplotypes of six SNPs and one CNV used in our previous work [6,14,16]. We found moderate correlations between the CYP2A6 activity metric and either PRSs of 0.40 ($p < 0.0001$), while the correlation between the Loukola *et al.* PRS and Baurley *et al.* PRS was 0.65 ($p < 0.0001$) (Supplementary Table 5).

Nicotine metabolism PRS quartiles significantly predict nicotine metabolism biomarkers (details in Table 1) (e.g., $\beta = 0.15$; 95% CI = 0.046–0.26 for Baurley PRS predicting NMR), and the variance explained by the PRS is modest (R^2 range: 9.2–16%; minimum $p = 7.6 \times 10^{-8}$) in the COGEND samples. Specifically, both nicotine metabolism PRSs predicted the most variance for the nicotine metabolism proportion cotinine/[cotinine + nicotine], than the NMR, and the least variance for the eNMR of experimentally ingested nicotine (Table 1). We obtained similar results using PRS as a continuous variable (Supplementary Table 6).

Nicotine metabolism GWAS top hit variant predicts substantial variance for nicotine metabolism biomarkers

The nicotine metabolism GWAS top hit variant (rs56113850) significantly predicted nicotine metabolism biomarkers (R^2 range: 14–17%; minimum $p = 4.4 \times 10^{-8}$) in the COGEND samples. Specifically, rs56113850 predicted the most variance for the NMR, then the eNMR of experimentally ingested nicotine, and the least variance for the nicotine metabolism proportion cotinine/[cotinine + nicotine] (Table 1).

Table 1. Polygenic risk scores based on nicotine metabolite ratio genome wide association study predict nicotine metabolism biomarkers.†												
Genetic predictor	Log NMR (n = 82)				Log Experimentally derived NMR (n = 184)				Log Metabolism proportion experimental (n = 184)			
	β	95% CI	p-value	R ²	β	95% CI	p-value	R ²	β	95% CI	p-value	R ²
Loukola PRS	0.15	0.046–0.26	6.0E-03	0.092	0.25	0.16–0.33	7.6E-08	0.14	0.047	0.028–0.065	2.3E-06	0.12
Baurley PRS	0.23	0.12–0.35	1.8E-04	0.16	0.25	0.16–0.35	4.4E-07	0.13	0.048	0.027–0.068	8.2E-06	0.10
Rs56113850	0.34	0.17–0.50	1.4E-04	0.17	0.40	0.26–0.54	4.4E-08	0.15	0.081	0.051–0.11	2.8E-07	0.14
† Linear regression models adjusted for age, sex and two principal components. R ² is variance explained by the specific PRS. PRS coded as quartiles. Biomarker outcomes are log transformed. NMR – ratio of cotinine : 3-hydroxycotinine. Experimentally derived NMR – ratio of deuterated cotinine : 3-hydroxycotinine 4 h after ingestion of deuterated nicotine. Metabolism proportion – ratio of deuterated cotinine (deuterated cotinine plus deuterated nicotine) 30 min after ingestion of deuterated nicotine. R2 reported here is associated with the predictor with the variance associated with other covariates (age, sex, PC2) deducted. NMR: Nicotine metabolite rate; PRS: Polygenic risk score.												

Table 2. Equivocal evidence for polygenic risk scores based on nicotine metabolite ratio genome wide association study predicting smoking quantity assessed by cigarettes per day.†												
Genetic predictor	UW-TTURC (n = 1031)			COGEND (n = 1866)			ARIC (n = 5027)			MESA (n = 1158)		
	β	95% CI	p-value	β	95% CI	p-value	β	95% CI	p-value	β	95% CI	p-value
Loukola PRS	0.032	-0.0086–0.072	0.12	0.019	-0.024–0.063	0.39	0.035	0.011–0.060	0.0051	0.014	-0.038–0.067	0.59
Baurley PRS	0.039	-0.0035–0.082	0.07	0.035	-0.011–0.080	0.14	0.021	-0.0023–0.043	0.077	-0.00058	-0.049–0.048	0.98
† Cigarettes per day (CPD), defined as a 4-level ordered trait (CPD ≤10; 11 ≤CPD ≤20; 21 ≤CPD ≤30; CPD ≥31, coded as 0, 1, 2, 3, respectively). Linear regression models on 4-level CPD, adjusted for age, sex, and two principal components. PRS coded as quartiles. ARIC: Atherosclerosis Risk in Communities; COGEND: Collaborative Genetic Study of Nicotine Dependence; MESA: Multi-Ethnic Study of Atherosclerosis; PRS: Polygenic risk score; UW-TTURC: University of Wisconsin Transdisciplinary Tobacco Use Research Center.												

Weak evidence for PRSs based on nicotine metabolism GWAS predicting smoking quantity

Nicotine metabolism PRSs (NMR PRS) based on the GWAS meta-analyses predict smoking quantity defined as cigarettes per day (CPD) in ARIC, but not in COGEND, MESA or UW-TTURC (Table 2). In ARIC, higher NMR PRS, which indicated faster metabolism, were associated with higher number of cigarettes smoked per day (Loukola *et al.* PRS β = 0.04; 95% CI: 0.01–0.06; p = 0.0051; Baurley *et al.* PRS β = 0.02; 95% CI: -0.002–0.04; p = 0.077). We reached similar results when modeling CPD as a continuous variable (Supplementary Table 7).

Given the null associations in the other three studies, we conducted power analyses to evaluate the detectable effect sizes given the sample sizes. For cigarettes per day (CPD) in UW-TTURC, COGEND and MESA, our sample had sufficient power (80%) to detect an association with an effect size of 0.078, 0.058 and 0.074 respectively. Therefore, we had power to detect only strong associations between the NMR PRS and smoking quantity.

No evidence for PRSs based on nicotine metabolism GWAS predicting smoking cessation

We found no evidence of association between nicotine metabolism PRSs based on the GWAS meta-analyses and smoking cessation defined as smoking abstinence at end of treatment and smoking relapse in the treatment trial (UW-TTURC), time to relapse in GERS subset of COGEND, and age of smoking cessation in community-based, cross-sectional studies (COGEND, ARIC, MESA), as shown in Table 3. We reached similar results of no association when modeling duration of smoking in these community studies. Similar results for different treatment conditions are included in the Supplementary Table 8.

We conducted power analyses to demonstrate the power regarding the findings of null associations. We had sufficient power (80%) to detect an association with an odds ratio of 0.71 for smoking abstinence in UW-TTURC. For age of quitting in COGEND, ARIC and MESA, our sample had sufficient power (80%) to detect an association with a hazard ratio of 1.25, 1.11 and 1.24, respectively. Therefore, we had power to detect only strong associations between the NMR PRS and smoking cessation.

Table 3. No evidence for polygenic risk scores based on nicotine metabolite ratio genome wide association study predicting smoking cessation.[†]

Genetic predictor	UW-TTUTC smokers with placebo [‡] (n = 130)			UW-TTUTC smokers with active medication [‡] (n = 937)			GERS time to relapse in all groups [§] (n = 479)			ARIC [¶] (n = 5054)			MESA [¶] (n = 1187)		
	OR	95% CI	p-value	OR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value
Loukola PRS	0.88	0.62–1.23	0.44	1.06	0.94–1.19	0.35	1.05	0.97–1.15	0.22	1.00	0.97–1.03	0.97	1.04	0.98–1.10	0.16
Baurley PRS	0.87	0.61–1.25	0.44	0.99	0.87–1.11	0.81	0.98	0.90–1.07	0.64	1.03	1.00–1.06	0.046	1.00	0.95–1.05	0.99

[†] PRS coded as quartiles.[‡] Logistic regression models on smoking abstinence, adjusted for age, sex and two principal components as covariates.[§] Cox regression models on time to smoking relapse, adjusted for age, sex and two principal components as covariates.[¶] Cox regression models on age of quitting, adjusted for age, sex and two principal components as covariates.

ARIC: Atherosclerosis Risk in Communities; GERS: Genetic and Environmental Risks for Smoking Characteristics and Cessation; HR: Hazard ratio; MESA: Multi-Ethnic Study of Atherosclerosis; OR: Odds ratio; PRS: Polygenic risk score; UW-TTUTC: University of Wisconsin Transdisciplinary Tobacco Use Research Center.

Discussion

This study examined whether PRSs derived from two large GWAS meta-analyses of nicotine metabolism measured by the NMR predict nicotine metabolism, smoking quantity and smoking cessation, when tested in independent samples. We used three community-based, cross-sectional studies and a randomized comparative effectiveness smoking cessation trial. We show that nicotine metabolism PRSs based on the GWAS meta-analyses significantly predict nicotine metabolism biomarkers (R^2 range: 6–15%; minimum $p = 1.1 \times 10^{-7}$). Further, use of the top hit variant predicts a substantial variance. This work shows that the PGRs derived in one sample can predict nicotine metabolism in other samples.

However, the predictive ability of nicotine metabolism PGRs for smoking quantity and smoking cessation is more modest. Though there is much evidence that nicotine metabolism predicts smoking behaviors and smoking cessation, translating this nicotine metabolism measures to PGRs reduces power and the signal is attenuated. Additional research is needed to develop a more comprehensive genetic risk score to predict smoking quantity and smoking cessation. Given that PRS could potentially increase prediction in other diseases [19], this finding of comparable variance explained by PRS and single top variant could be explained by the use of many more variants including variants below the GWAS threshold is often needed to maximize the predictive power; the size of subjects in the two discovery GWAS are modest, limiting their power to detect biologically significant variant for nicotine metabolism.

These findings extend the existing research on genetic variation in *CYP2A6* and nicotine metabolism. The NMR is a biomarker of nicotine metabolism that reflects both genetic and environmental influences on nicotine metabolism and clearance [35]. The PRSs of nicotine metabolism capture variation in *CYP2A6*, which in turn influences nicotine metabolism. Because nicotine metabolism is related to smoking quantity and smoking cessation, our goal was to test if these PRSs of NMR can be used to predict smoking quantity and smoking cessation.

Figure 1 summarizes the existing evidence regarding how genetic variation in nicotine metabolism (as captured by different versions of PRSs) related to nicotine metabolism biomarkers, smoking quantity and smoking cessation [1,4,6,12,16–18,36–42]. The effect of nicotine metabolism genes as captured by the previously reported *CYP2A6* activity metric (based on customized genotyping and haplotypes using six SNPs and one CNV) captured approximately 70% of variance for the metabolism proportion (a metabolism marker different than NMR), 10–18% of variance for the NMR and predicted smoking cessation in a smoking cessation trial [16]. In contrast, the effect of nicotine metabolism genes as captured by NMR PRS (based on GWAS meta-analyses of NMR) captured about 10–12% of variance of the metabolism proportion, 9–16% of variance of the NMR, and did not significantly predict smoking cessation in our study. The effect of nicotine metabolism genes as captured by the top variant rs56113850 captured about 14% of variance of the metabolism proportion, 15–17% of variance of the NMR, and did not significantly predict smoking cessation. These differences likely result from many factors including the use of different metabolism biomarkers (NMR vs metabolism proportion), use of single SNP versus a PRS and weights based on discovery GWAS of single SNP analyses versus haplotype analyses, which could crucially affect the validity and utility of these nicotine metabolism scores. In addition, recent prospective evidence suggests that the effect of NMR on smoking cessation success could be complex [13].

Figure 1. Synthesis of literature on nicotine metabolism genes and smoking behaviors.

CPD: Cigarettes per day.

This study has several limitations. First, smoking cessation in the community samples was self-reported and not assessed using biochemical confirmation. Further, self-reported age of quitting in the cross-sectional studies, instead of prospective studies, were used to estimate true age of quitting, so this could suffer from subjects' recall bias. However, abstinence was biochemically confirmed in the effectiveness trial. Second, we could not test the effect of smoking cessation treatment in the community samples because this information was not available. We assume the use of medication is not common because the majority of smokers quit without medication in the general population [44]. Third, this work assumed additive effects between GWAS variants instead of other nonadditive or interactive effects, which may optimize their predictive validity. Fourth, we may not have captured other important genetic variation in the *CYP2A6* region that could have contributed to the results. We have tested PRS based on GWAS level significant variants instead of including less significant yet potentially informative variants. The *CYP2A6* gene is highly polymorphic [45], including many variants in Europeans [5]. The complex genetic architecture in this chromosomal region and other metabolic pathways in addition to nicotine metabolism could play a role. Further, the mechanisms underlying nicotine metabolism genes such as *CYP2A6* and smoking cessation are not entirely clear. Further, the generalizability of these results is limited given the inclusion of only individuals of European Ancestry. The populations from which the two discovery GWASs were derived from differed. Finally, the statistical power of these studies is limited due to the sample size. A meta-analysis of more studies with larger sample sizes is required to determine the utility of these PRSs in clinical care and public health.

Conclusion

Keeping in mind the above limitations, this study extends previous work on nicotine metabolism genes (e.g., *CYP2A6*) and cessation to reach the following conclusions: 1) Results suggest that nicotine metabolism PRSs based on GWAS meta-analyses may be a useful predictor of an individual's nicotine metabolism; 2) The top variant in GWAS of NMR explained a substantial variance of nicotine metabolism markers; 3) Additional research is needed to develop a more comprehensive genetic risk score that models nicotine metabolism and predicts smoking quantity and smoking cessation. Many other genes have been nominated as predictive of smoking quantity and cessation [46], and we anticipate more genes will be identified as playing a role in determining smoking behaviors. A

more comprehensive risk prediction model that incorporates multiple genetic markers and nongenetic predictors will lay the foundation for a personalized smoking cessation treatment algorithm [47].

Future perspective

Nicotine metabolism PRSs may be a useful predictor of an individual's nicotine metabolism, and possibly smoking cessation. Many other genes may also predict smoking cessation. Risk prediction modeling to incorporate multiple genetic markers and nongenetic predictors will lay the foundation for a personalized treatment algorithm. Application of these PRSs in large datasets (e.g., UK Biobank, 23 and Me, Million Veteran Project) in predicting smoking behaviors and related disorders would serve as great opportunities to evaluate their clinical and public health impact.

Summary points

Aim

- Variation in the gene that encodes the primary nicotine metabolism enzyme, cytochrome P450 2A6 (*CYP2A6*), is possibly associated with smoking behaviors.
- This study examined whether polygenic risk scores (PRSs) for nicotine metabolism predict smoking quantity, smoking cessation and nicotine metabolism in independent datasets.

Materials & methods

- PRSs for nicotine metabolism derived from two large genome wide association studies (GWAS) meta-analyses were computed in three community-based, cross-sectional studies and one randomized, comparative effectiveness smoking cessation trial.

Results

- Nicotine metabolism PRSs based on the GWAS meta-analyses significantly predicted nicotine metabolism biomarkers.
- Evidence was equivocal regarding the use of nicotine metabolism PRSs in predicting smoking quantity.
- Nicotine metabolism score was not significantly associated with smoking cessation.

Conclusion

- Nicotine metabolism PRSs based on GWAS meta-analyses may be a useful predictor of an individual's nicotine metabolism.
- These metabolism scores require additional refinement to predict smoking quantity and smoking cessation, and additional research is needed to develop a more comprehensive genetic risk score leveraging biology and association.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at:

<https://www.futuremedicine.com/doi/suppl/10.2217/pgs-2018-0081>

Acknowledgements

The authors thank S Fisher, L Fox and N Smock for editorial and data support of the manuscript.

Financial & competing interests disclosure

This research was supported by R01DA036583, P30CA091842, U01HG004422 (LJB), R01DA038076, P30 CA091842-16S2, U19CA203654, K08DA030398, and KL2RR024994 (LSC), P50CA84724, K05CA139871, P50DA19706 (TBB), R01DA026911 (NLS).

The Collaborative Genetic Study of Nicotine Dependence (COGEND; PI: L Bierut) was supported by grant P01CA089392 from the National Cancer Institute. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the NIH to The Johns Hopkins University, contract number HHSN268200782096.

The Atherosclerosis Risk in Communities Study (ARIC; PI: E Boerwinkle) was obtained from dbGaP through study accession number phs000090. ARIC is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and NIH contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the NIH and NIH Roadmap for Medical Research.

MESA and the MESA SHARE project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC95166, N01-HC-95167, N01-HC-95168, N01-HC-95169 and CTSA UL1-RR-024156. Funding for SHARE genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (CA, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetric Genome-Wide Human SNP Array 6.0.

Genotyping services for the UW-TTURC sample were provided by the Center for Inherited Disease Research (CIDR) and DNA extraction supported by the Wisconsin State Laboratory of Hygiene. Funding support for CIDR was provided by NIH grant U01HG004438 and NIH contract HHSN268200782096C to The Johns Hopkins University. Assistance with genotype cleaning was provided by the Gene Environment Association Studies (GENEVA) Coordinating Center (U01 HG004446).

Glaxo Wellcome provided bupropion at no cost in the UW-TTURC clinical trial. The Wisconsin State Laboratory of Hygiene provided considerable technical assistance in this research effort in the form of DNA extraction, and this research was also aided by the Wisconsin Partnership Program. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Conflicts of interest

LJ Bierut is listed as an inventor on issued US patent 8,080,371, 'Markers for Addiction' covering the use of certain SNPs in determining the diagnosis, prognosis and treatment of addiction. N Saccone is the spouse of S Saccone who is also listed as an inventor on the above patent. All other authors declare no potential conflict of interest.

References

Papers of special note have been highlighted as: • of interest •• of considerable interest

1. Bloom AJ, Harari O, Martinez M *et al.* Use of a predictive model derived from *in vivo* endophenotype measurements to demonstrate associations with a complex locus, CYP2A6. *Hum. Mol. Genet.* 21(13), 3050–3062 (2012).
- **CYP2A6 activity metrics predict one of nicotine metabolism biomarker, nicotine metabolism proportion.**
2. Fujieda M, Yamazaki H, Saito T *et al.* Evaluation of CYP2A6 genetic polymorphisms as determinants of smoking behavior and tobacco-related lung cancer risk in male Japanese smokers. *Carcinogenesis* 25(12), 2451–2458 (2004).
3. Schoedel KA, Hoffmann EB, Rao Y, Sellers EM, Tyndale RF. Ethnic variation in CYP2A6 and association of genetically slow nicotine metabolism and smoking in adult Caucasians. *Pharmacogenetics* 14(9), 615–626 (2004).
4. Thorgeirsson TE, Gudbjartsson DF, Surakka I *et al.* Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat. Genet.* 42(5), 448–453 (2010).
5. Haberl M, Anwald B, Klein K *et al.* Three haplotypes associated with CYP2A6 phenotypes in Caucasians. *Pharmacogenet. Genomics* 15(9), 609–624 (2005).
6. Bloom J, Hinrichs AL, Wang JC *et al.* The contribution of common CYP2A6 alleles to variation in nicotine metabolism among European Americans. *Pharmacogenet. Genomics* 21(7), 403–416 (2011).
7. Lerman C, Jepson C, Wileyto EP *et al.* Genetic variation in nicotine metabolism predicts the efficacy of extended-duration transdermal nicotine therapy. *Clin. Pharmacol. Ther.* 87(5), 553–557 (2010).
8. Patterson F, Schnoll RA, Wileyto EP *et al.* Toward personalized therapy for smoking cessation: a randomized placebo-controlled trial of bupropion. *Clin. Pharmacol. Ther.* 84(3), 320–325 (2008).
9. Schnoll RA, Patterson F, Wileyto EP, Tyndale RF, Benowitz N, Lerman C. Nicotine metabolic rate predicts successful smoking cessation with transdermal nicotine: a validation study. *Pharmacol. Biochem. Behav.* 92(1), 6–11 (2009).
10. Benowitz NL, Pomerleau OF, Pomerleau CS, Jacob P 3rd. Nicotine metabolite ratio as a predictor of cigarette consumption. *Nicotine Tob. Res.* 5(5), 621–624 (2003).
11. Ray R, Tyndale RF, Lerman C. Nicotine dependence pharmacogenetics: role of genetic variation in nicotine-metabolizing enzymes. *J. Neurogenet.* 23(3), 252–261 (2009).
12. Lerman C, Schnoll RA, Hawk LW Jr *et al.* Use of the nicotine metabolite ratio as a genetically informed biomarker of response to nicotine patch or varenicline for smoking cessation: a randomised, double-blind placebo-controlled trial. *Lancet Respir Med.* 3(2), 131–138 (2015).
13. Fix BV, O'Connor RJ, Benowitz N *et al.* Nicotine metabolite ratio (NMR) prospectively predicts smoking relapse: longitudinal findings from ITC surveys in five countries. *Nicotine Tob Res.* 19(9), 1040–1047 (2017).
14. Bloom AJ, Baker TB, Chen LS *et al.* Variants in two adjacent genes, EGLN2 and CYP2A6, influence smoking behavior related to disease risk via different mechanisms. *Hum. Mol. Genet.* 23(2), 555–561 (2013).

15. Bloom AJ, Hartz SM, Baker TB *et al.* Beyond cigarettes per day. A genome-wide association study of the biomarker carbon monoxide. *Ann. Am. Thoracic Soc.* 11(7), 1003–1010 (2014).
16. Chen LS, Bloom AJ, Baker TB *et al.* Pharmacotherapy effects on smoking cessation vary with nicotine metabolism gene (*CYP2A6*). *Addiction* 109(1), 128–137 (2014).
- **CYP2A6 activity metrics predict smoking cessation and response to nicotine replacement therapy.**
17. Loukola A, Buchwald J, Gupta R *et al.* A genome-wide association study of a biomarker of nicotine metabolism. *PLoS Genet.* 11(9), e1005498 (2015).
- **Genome wide association studies meta-analysis of nicotine metabolite ratio.**
18. Baurley JW, Edlund CK, Pardamean CI *et al.* Genome-wide association of the laboratory-based nicotine metabolite ratio in three ancestries. *Nicotine Tob. Res.* 18(9), 1837–1844 (2016).
- **Genome wide association studies meta-analysis of nicotine metabolite ratio.**
19. Khera AV, Chaffin M. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet.* 50(9), 1219–1224 (2018).
20. Swan GE, Benowitz NL, Lessov CN, Jacob P 3rd, Tyndale RF, Wilhelmsen K. Nicotine metabolism: the impact of *CYP2A6* on estimates of additive genetic influence. *Pharmacogenet. Genomics* 15(2), 115–125 (2005).
21. Saccone NL, Saccone SF, Hinrichs AL *et al.* 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(4), 453–466 (2009).
22. Saccone NL, Wang JC, Breslau N *et al.* The *CHRNA5-CHRNA3-CHRNA4* nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res.* 69(17), 6848–6856 (2009).
23. Piper ME, Smith SS, Schlam TR *et al.* A randomized placebo-controlled clinical trial of 5 smoking cessation pharmacotherapies. *Arch. Gen. Psychiatry* 66(11), 1253–1262 (2009).
24. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The fagerstrom test for nicotine dependence: a revision of the fagerstrom tolerance questionnaire. *Br. J. Addict.* 86(9), 1119–1127 (1991).
25. Stephens SH, Hartz SM, Hoft NR *et al.* Distinct loci in the *CHRNA5/CHRNA3/CHRNA4* gene cluster are associated with onset of regular smoking. *Genet. Epidemiol.* 37(8), 846–859 (2013).
26. Piper ME, Federman EB, McCarthy DE *et al.* Efficacy of bupropion alone and in combination with nicotine gum. *Nicotine Tob. Res.* 9(9), 947–954 (2007).
27. Sobell LC, Sobell MB. Timeline follow-back: a technique for assessing self-reported alcohol consumption. In: *Measuring Alcohol Consumption: Psychosocial and Biological Methods* Litten RZ, Allen JP. (Eds.). Humana Press, Totowa, NJ, US, 41–72 (1992).
28. Johnson EO, Hancock DB, Levy JL *et al.* Imputation across genotyping arrays for genome-wide association studies: assessment of bias and a correction strategy. *Hum. Genet.* 132(5), 509–522 (2013).
29. 1000 Genomes Project Consortium; Abecasis GR, Auton A *et al.* An integrated map of genetic variation from 1092 human genomes. *Nature* 491(7422), 56–65 (2012).
30. Hartz SM, Horton AC, Oehlert M *et al.* Association between substance use disorder and polygenic liability to schizophrenia. *Biol. Psychiatry* 82(10), 709–715 (2017).
31. Carey CE, Agrawal A, Bucholz KK *et al.* Associations between polygenic risk for psychiatric disorders and substance involvement. *Front. Genet.* 7, 149 (2016).
32. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. *Bioinformatics* 31(9), 1466–1468 (2015).
33. Purcell S, Neale B, Todd-Brown K *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81(3), 559–575 (2007).
34. Chang SH, Luo S, O'Brian KK *et al.* Association between metformin use and transformation of monoclonal gammopathy of undetermined significance to multiple myeloma in U.S. veterans with diabetes mellitus: a population-based cohort study. *Lancet Haematol.* 2(1), e30–e36 (2015).
35. St Helen G, Novalen M, Heitjan DF *et al.* Reproducibility of the nicotine metabolite ratio in cigarette smokers. *Cancer Epidemiol. Biomark. Prev.* 21(7), 1105–1114 (2012).
36. Schuit E, Panagiotou OA, Munafo MR, Bennett DA, Bergen AW, David SP. Pharmacotherapy for smoking cessation: effects by subgroup defined by genetically informed biomarkers. *Cochrane Database Syst. Rev.* 9, CD011823 (2017).
37. Ho MK, Mwenifumbo JC, Al Koudsi N *et al.* Association of nicotine metabolite ratio and *CYP2A6* genotype with smoking cessation treatment in African-American light smokers. *Clin. Pharmacol. Ther.* 85(6), 635–643 (2009).
38. Liu T, David SP, Tyndale RF *et al.* Associations of *CYP2A6* genotype with smoking behaviors in southern China. *Addiction* 106(5), 985–994 (2011).
39. Liu JZ, Tozzi F, Waterworth DM *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.* 42(5), 436–440 (2010).

40. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* 42(5), 441–447 (2010).
41. Chen LS, Baker TB, Piper ME *et al.* Interplay of genetic risk factors (CHRNA5-CHRNA3-CHRNA4) and cessation treatments in smoking cessation success. *Am. J. Psychiatry* 169(7), 735–742 (2012).
42. Chen LS, Hung RJ, Baker T *et al.* CHRNA5 risk variant predicts delayed smoking cessation and earlier lung cancer diagnosis – a meta-analysis. *J. Natl Cancer Inst.* 107(5), djv100 (2015).
43. Johnstone E, Benowitz N, Cargill A *et al.* Determinants of the rate of nicotine metabolism and effects on smoking behavior. *Clin. Pharmacol. Ther.* 80(4), 319–330 (2006).
44. Hughes JR, Keely J, Naud S. Shape of the relapse curve and long-term abstinence among untreated smokers. *Addiction* 99(1), 29–38 (2004).
45. McDonagh EM, Wassenaar C, David SP *et al.* PharmGKB summary: very important pharmacogene information for cytochrome P-450, family 2, subfamily A, polypeptide 6. *Pharmacogenet. Genomics* 22(9), 695–708 (2012).
46. Uhl GR, Liu QR, Drgon T *et al.* Molecular genetics of successful smoking cessation: convergent genome-wide association study results. *Arch. Gen. Psychiatry* 65(6), 683–693 (2008).
47. Lotrich FE. The emerging potential of pharmacogenetics in psychiatry. *Am. J. Psychiatry* 169(7), 681–683 (2012).