Published in final edited form as:

Drug Alcohol Depend. 2016 October 1; 167: 95–102. doi:10.1016/j.drugalcdep.2016.07.029.

Polygenic Risk Scores For Cigarettes Smoked Per Day Do Not Generalize To A Native American Population

Jacqueline M. Otto[a], **Ian R. Gizer**[a], **Chris Bizon**[b], **Kirk C. Wilhelmsen**[b,c], and **Cindy L. Ehlers**[d,*]

[a]Department of Psychological Science, University of Missouri, 210 McAlester Hall, Columbia, MO 65211

[b]Renaissance Computing Institute (RENCI), 100 Europa Drive, Suite 540, Chapel Hill, NC 27517

[c]Departments of Genetics and Neurology, University of North Carolina at Chapel Hill, 120 Mason Farm Road, 5093 Genetic Medicine Building, CB#7264, Chapel Hill, NC 27599

[d]Department of Molecular and Cellular Neurosciences, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Abstract

Background—Recent studies have demonstrated the utility of polygenic risk scores (PRSs) for exploring the genetic etiology of psychiatric phenotypes and the genetic correlations between them. To date, these studies have been conducted almost exclusively using participants of European ancestry, and thus, there is a need for similar studies conducted in other ancestral populations. However, given that the predictive ability of PRSs are sensitive to differences in linkage disequilibrium (LD) patterns and minor allele frequencies across discovery and target samples, the applicability of PRSs developed in European ancestry samples to other ancestral populations has yet to be determined. Therefore, the current study derived PRSs for cigarettes per day (CPD) from predominantly European-ancestry samples and examined their ability to predict nicotine dependence (ND) in a Native American (NA) population sample.

Method—Results from the Tobacco and Genetics Consortium's meta-analysis of genome-wide association studies of CPD were used to compute PRSs in a NA community sample (*N*=288). These scores were then used to predict ND diagnostic status.

*Corresponding author at: Cindy L. Ehlers, 10550 N. Torrey Pines Road, SP30-1501, La Jolla, CA 92037. Telephone (858) 784-7058.

Contributors

Jacqueline M. Otto (JMO), Ian R. Gizer (IRG), Chris Bizon (CB), Kirk C. Wilhelmsen (KCW), Cindy L. Ehlers (CLE)
Study conception and design: Otto, Gizer
Acquisition of data: Ehlers, Wilhelmsen
Analysis and interpretation of data: Otto, Gizer, Bizon
Drafting of manuscript: Otto, Gizer, Bizon, Ehlers, Wilhelmsen
All authors have read the manuscript and approve of its submission to Drug and Alcohol Dependence.

Conflict of Interest

No conflict declared

Results—The PRS was not significantly associated with liability for ND in the full sample. However, a significant interaction between PRS and percent NA ancestry was observed. Risk scores were positively associated with liability for ND at higher levels of European ancestry, but no association was observed at higher levels of NA ancestry.

Conclusion—These findings illustrate how differences in patterns of LD across discovery and target samples can reduce the predictive ability of PRSs for complex traits.

Keywords

smoking; polygenic; nicotine dependence; genetic association; American Indian

1. INTRODUCTION

Tobacco use and nicotine dependence (ND) remain highly prevalent in the United States. Approximately 16.8% of the U.S. adult population self-identified as current smokers in 2014, and the majority of these individuals reported smoking at least once daily (Centers for Disease Control and Prevention [CDC] 2015). Estimates from nationally representative samples of past-year daily smokers indicate that over 60% meet criteria for ND (Donny and Dierker, 2007). Continued use of tobacco is the leading cause of preventable disease and mortality in the U.S. (U.S. Department of Health and Human Services [USDHHS], 2014), and thus represents a significant public health concern (USDHHS, 2010).

Of additional note, epidemiological studies suggest rates of tobacco use and ND differ across racial and ethnic groups, indicating that this serious public health issue has a disproportionate effect on certain minority groups. Smoking rates in the U.S. are highest among American Indian and Alaska Native populations; in 2014, 29.2% of non-Hispanic American Indian and Alaska Native individuals smoked cigarettes versus 16.8% in the U.S. overall (CDC, 2015). Although prevalence rates differ across tribes within these populations, and many sociocultural and economic factors influence these prevalence rates, the high proportion of American Indian individuals who report smoking puts this group at a uniquely high risk for disease and related negative consequences (USDHHS, 1998).

Quantitative genetics studies estimate that a substantial proportion of variance in the diagnosis of nicotine dependence and smoking behavior is due to genetic factors. At a population level, twin and family studies have indicated that genetic factors account for approximately 44% of the variance in smoking initiation (Vink et al., 2005), 50% of the variance in smoking quantity (Mackillop et al., 2010), and 33–75% of the variance in the liability for ND (Agrawal and Lynskey, 2008; Agrawal et al., 2012). In Native American (NA) populations, the heritability estimates (h^2) for regular and persistent tobacco use approach similar levels and range from 0.37–0.53 and 0.34–0.46, respectively (Ehlers and Wilhelmsen, 2006). As a result, published evidence suggests that these phenotypes are similarly heritable across European and NA populations (Ehlers and Gizer, 2013).

Evidence from meta-analyses of genome-wide association studies (GWASs) have identified common variants associated with number of cigarettes smoked per day (CPD), including those within the nAChR genes on chromosome 15q24-q25 (Chen et al., 2012; David et al.,

2012; Liu et al., 2010; Saccone et al., 2010; Thorgeirsson et al., 2010; Tobacco and Genetics Consortium, 2010; Ware et al., 2011), and variants in regions that contain *CYP2A6* and *CYP2B6* on chromosome 9q13 and *CHRNB3* and *CHRNA6* on chromosome 8p11 (Thorgeirsson et al., 2010). Though most of the genome-wide approaches have focused on quantitative measures of smoking behavior, ND has also been associated with variants within the nAChR genes (Rice et al., 2012; Thorgeirsson et al., 2008) and may account for the robust relationships observed between higher levels of self-reported cigarette smoking and variants within this region (Rice et al., 2012). Despite these initial successes, it is important to note that these efforts have focused almost exclusively on participants of European ancestry, thus limiting the generalizations that can be drawn from these studies in relation to other ancestral groups, including NA populations.

It is also notable that the amount of variance explained by each SNP identified in the described studies fails to approach the heritability estimates reported by quantitative genetics studies. This is because complex traits such as smoking behaviors and ND are thought to be highly polygenic in nature with hundreds of variants, each of small individual effect (e.g., R^2<0.005) contributing to the development of the phenotype. Therefore, modeling the additive or cumulative effects of associated variants has the potential to explain a higher proportion of variation in a trait relative to a single variant. To this effect, methods were developed to calculate and test the association of polygenic risk scores (PRSs) with a phenotype of interest (The International Schizophrenia Consortium [ISC], 2009; Wray et al., 2007). Notably, GWASs of common variants suggest that many, if not most, of the variants involved in complex trait etiology have too small an effect size to reach genome-wide significance levels even when sample sizes are large (~40,000). Thus, PRSs are typically created by selecting variants that achieve a pre-designated significance level (e.g., p<0.01 rather than $p<5 \times 10^{-8}$) from a large-scale ‘discovery’ GWAS with the assumption that their cumulative effect will reduce the influence of spurious associations. Once variants are selected, the number of risk alleles at each marker are tallied, weighted by their respective regression beta weight or log odds ratio, and summed to create a risk score for each individual in an independent replication sample. These risk scores are then used to determine the proportion of variation in the trait that can be explained by their cumulative effects.

The application of PRSs to the study of psychiatric phenotypes has been met with a fair amount of success. For example, PRSs were able to explain 6% of the variation in schizophrenia diagnostic rates, a considerable improvement compared to the variation explained by individual variants, which was less than 1% (Ripke et al., 2011). Other psychiatric phenotypes (e.g., de Moor et al., 2015; de Zeeuw et al., 2014; Hamshere et al., 2013a; 2013b; ISC, 2009; Ruderfer et al., 2013; Sklar et al., 2011) have demonstrated similar results using cumulative measures of genetic risk, including tobacco and other substance use phenotypes (Salvatore et al., 2014). PRSs for variants associated with CPD in the Tobacco and Genetics Consortium (TAG) meta-analysis (Tobacco and Genetics Consortium [TAG], 2010) significantly predicted tobacco use at ages 20 and 24 (Vrieze et al., 2012), and in another study, explained a small but significant proportion of the variation in number of drinks consumed per week (0.4–0.5%) and age of cannabis initiation (0.6–0.9%; Vink et al., 2014). These findings demonstrate the potential of PRS approaches to further our understanding of how genetic influences contribute to risk for individual

substance use phenotypes, as well as how these genetic influences might contribute to shared risk across substances.

Despite these promises, limitations exist in the application and interpretation of findings from studies employing PRS methods, including aspects of the discovery sample. Adequate sample size is necessary for precise score estimation in the discovery population, and the optimal *p*-value threshold used for selecting score variants from the initial genome-wide analysis depends on the size of the discovery sample. With adequately powered discovery samples, one can lower the threshold for variant inclusion, as these are more likely to reflect true positive associations with the phenotype (Wray et al., 2014). Additionally, disparate patterns of linkage disequilibrium (LD) and differences in marker allele frequencies between discovery and target samples are thought to attenuate effects in PRS analyses (ISC, 2009; Wray et al., 2007).

This last consideration has important implications. Population stratification and differential patterns of LD across racial and ethnic groups are well-known confounds that can bias results from genetic association studies (Price et al., 2006). The vast majority of GWASs are conducted with individuals of European descent, and given that methods for computing PRSs depend on summary statistics from GWASs, risk alleles identified from these studies may be specific to that ancestral group or include tag single nucleotide polymorphisms (SNPs) not found in other populations (Domingue et al., 2014). In addition, a standard practice in the construction of PRSs involves the LD-based pruning of GWAS results, so that only a single SNP from a given genomic region is included in the PRS. This is accomplished by selecting an initial set of SNPs that meet a predefined *p*-value, and then, beginning with the most highly associated SNP and proceeding in order of descending strength of association, using LD statistics to iteratively remove correlated markers from the set (e.g., R^2<0.2 across 500kb; Wray et al., 2014). This procedure ensures that only independent association signals are included, and thus, protects against overestimating the explanatory power of the PRS (ISC, 2009; Wray et al., 2007).

The existing literature on PRSs for psychiatric phenotypes has been almost exclusively restricted to discovery and target samples composed of individuals of European descent. Because LD-based clumping procedures are employed in order to restrict the PRS to a set of independent signals, one disadvantage to this approach is the reliance on a single SNP to effectively tag a region based on its LD with other nearby markers. The extent to which these findings would replicate in a sample with different patterns of LD and minor allele frequencies is unknown. Given epidemiological data indicating that NA populations have the highest rates of smoking and ND of any U.S. ethnic group and the heritability of these traits in this population (Wilhelmsen and Ehlers, 2005), validating the predictive ability of PRSs for smoking phenotypes in this population is important for understanding the genetic contributions to these behaviors. In addition, the availability of larger consortium datasets mainly composed of European-ancestry individuals highlights a growing need to address disparities in the racial and ethnic populations represented in the psychiatric genetics literature. Thus, the current study aimed to evaluate whether a PRS derived from a large-scale GWAS of CPD conducted in a predominantly European-ancestry sample would generalize to, and thus predict tobacco use phenotypes, in a NA population sample.

2. METHODS

2.1. Discovery Sample from Tobacco and Genetics (TAG) Consortium

The TAG consortium's GWAS meta-analysis of CPD was composed of European-ancestry individuals (*N*=38,131) drawn from 16 studies (TAG, 2010). Participants who endorsed smoking 100 cigarettes in their lifetime were asked to report on either average or maximum CPD. Each of the 16 studies performed their own genotyping, quality control, and imputation. Further details on the methods and summary statistics for the approximately 2.5 million markers in common across the 16 studies are reported in the published meta-analysis and available for download (http://www.nature.com/ng/journal/v42/n5/full/ng.571.html; TAG, 2010).

2.2. Target Sample from Native American (NA) Community Population

Participants were recruited from eight geographically contiguous reservations and eligible for participation if they reported at least 1/16th NA heritage, and were between 18 and 82 years old. The full sample included 775 participants nested within 161 families ranging in size from 1 to 319 individuals. Quality control procedures excluded 78 individuals from analyses. Sixty-seven samples could not be sequenced due to insufficient or low-quality deoxyribonucleic acid (DNA), and 11 samples could not be successfully identified when matching kinship coefficients to self-reported pedigree structure. Of the 697 individuals remaining, 288 individuals with available sequence and smoking phenotype data were selected for the current study. At least 50% NA heritage was reported by 39.6% of participants (median percent NA heritage=44.4), based on federal Indian blood quantum. Average age of the final sample was 32.7 years (*SD*=14.9), and predominantly female (55.6%; *N*=160). ND was reported by 37.5% of participants (*N*=108), and the average CPD was 14.1 (*SD*=12.7).

2.3. Measures

The Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) was administered to assess for CPD and DSM-IV ND, as well as demographic information. The SSAGA is an empirically validated semi-structured psychiatric interview with high test-retest reliability (kappas=0.70–0.90) for specific substance dependence diagnoses (Bucholz et al., 1994), and has been used successfully in studies with NA populations (Wall et al., 2003). The SSAGA uses a screening question asking whether participants have smoked 100 cigarettes in their lifetime. Participants who respond negatively are designated as non-smokers, not administered the full tobacco use section, and classified as not meeting criteria for ND. Thus, while ND diagnoses could be assigned to all participants, regardless of smoking history, CPD was only assessed in the limited number of individuals who endorsed smoking at least 100 cigarettes in their lifetime. Nonetheless, previous studies have reported substantial genetic correlations between ND and CPD (Vink et al., 2009), and thus, the ND diagnosis (coded 0/1) was used as the primary phenotype in the present study.

2.4. Sequencing

Blood derived DNA was sequenced using Illumina low-coverage whole genome sequencing using HiSeq2000 sequencers (Illumina, San Diego, CA). Approximately 80% of the samples were sequenced at a coverage depth between 3X and 12X (full range of coverage depth: 1X - 31X) with coverage depth evenly distributed across the genome. Sequence reads were aligned using blocked multiple-sequence alignment (BMA), and realigned near indels with the Genome Analysis Toolkit (GATK). Variants were called using the LD-aware variant caller Thunder (Li, 2011) to increase the accuracy of the variant calls. Variant call quality was assessed through comparison of the sequencing results to exome array genotypes (Affymetrix Exome 1A microarray) for all subjects, resulting in a 97.5% concordance rate. The general methodology for the sequencing and advantages of using the LD-aware variant calls in this sample have been reported previously (Bizon et al., 2014).

The TAG consortium meta-analytic data corresponded to build 36 (hg18) of the NCBI reference assembly (TAG, 2010), and chromosomal coordinates for these SNPs were converted to assembly build NCBI37 (hg19) by using the UCSC liftOver tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver). The appropriate tables provided by dbSNP were used to confirm strand orientation across builds, as well as identify and merge variants with multiple dbSNP rsIDs (RsMergeArch; http://www.ncbi.nlm.nih.gov/SNP/snp_db_table_description.cgi?t=RsMergeArch) or remove variants identified as having mapping errors (SNPChrPosOnRef: http://www.ncbi.nlm.nih.gov/SNP/snp_db_table_description.cgi?t=SNPChrPosOnRef). Concordance of strand alignment and allele coding was then confirmed by comparing the hg19 liftOver results for the non-ambiguous TAG consortium SNPs (i.e., excluding A/T and C/G SNPs) with the European cohort of the 1000 Genomes Project (1KGP) and with the NA target sample whole-genome sequence data. This comparison did not identify any errors in strand alignment or allele codings across datasets, indicating the conversion from build hg18 to hg19 was successful. Thus, all SNPs from the TAG consortium GWAS meta-analysis (including A/T and C/G SNPs) were retained for PRS construction.

2.5. Ancestry estimations

Ancestry proportions in the sample were analyzed using a supervised clustering approach that combined the algorithm implemented in the ADMIXTURE software (Alexander et al., 2009) in conjunction with a reference panel containing genotype information at about 300k strand-unambiguous SNPs. The ancestry estimates were then further refined through a noise reduction approach via bootstrapping (Libiger and Schork, 2012) and identified as corresponding to the four major continental populations: African, East Asian, European, and NA.

2.6. Data Analysis

The data analysis process was composed of three steps: (1) calculate LD statistics for SNPs included in the TAG meta-analysis using genotype data from the European cohort of the 1KGP (The 1000 Genomes Project [1KGP] Consortium, 2012), and use these statistics to create a set of independent markers, (2) match resulting independent TAG SNPs with NA sequence data, and (3) calculate CPD-based PRSs in the NA sample and conduct association

analyses to predict ND using linear mixed models in the lrgpr software package (Hoffman et al., 2014) as described below.

The initial TAG dataset contained 2,459,119 markers with summary statistics, including *p*-values, beta coefficients, and risk alleles for each marker. After removing markers with *p*-values >0.50, the remaining 1,257,959 SNPs were matched to the European-ancestry cohort of the 1KGP (1KGP Consortium, 2012), yielding 1,244,893 SNPs. Genotype data for these SNPs from the 1KGP were used along with the *p*-values from the TAG consortium GWAS to conduct LD-based pruning in PLINK 1.07 in order to extract the most highly significant, independent TAG markers from groups of correlated SNPs. This pruning procedure was conducted with seven different significance thresholds for index SNPs, ranging from $p \leq 0.01$ to $p \leq 0.5$. In this way, the most highly significant independent SNPs with a *p*-value less than or equal to each of the specified LD thresholds were identified and selected for the calculation of PRSs.

The resulting dataset of TAG SNPs following the LD-based clumping procedures contained 255,762 SNPs when the *p*-value threshold was set to ≤ 0.5; matches for these SNPs in the NA sample were identified, yielding a final dataset of 240,040 SNPs. PRSs were then generated using these markers in the NA sample for each of the seven significance thresholds used to run the LD-based clumping procedures. Individual PRSs were calculated by summing the number of TAG risk alleles (0, 1, or 2) at each marker included in the significance threshold, weighted by the SNP's regression coefficient from the TAG meta-analysis. These seven sets of PRSs were then used in family-based genetic association analyses in lrgpr (Hoffman et al., 2014) to predict ND in the NA target sample. Similar to many packages that conduct genetic association analyses, lrgpr uses a linear mixed model approach that includes the pairwise genetic similarity between all participant pairs approximated from genotyped markers as a random effect to account for population structure and genetic relatedness. Covariates included sex, age, age-squared, and ancestry estimates. The main effects of PRS and percent NA ancestry were tested first for association with liability for ND. Subsequent models included the interaction of PRS and percent NA ancestry in order to evaluate whether NA ancestry moderated the effects of PRS on liability for ND.

3. RESULTS

The variance in ND explained by the seven sets of risk scores ranged from 0.0–0.3% (see Table 1 for complete results). Notably, the main effect of the PRS created using a *p*-value inclusion threshold of 0.01 on ND was qualified by a significant interaction with percent NA ancestry (p=0.019; Table 2), and this model accounted for 2.2% of the variation in liability for ND, which is similar in magnitude to previous reports investigating PRSs for smoking phenotypes (Meyers et al., 2013; Vink et al., 2014). In order to evaluate the direction of effect for this interaction, secondary analyses were conducted on the top and bottom thirds of the sample after partitioning on percent NA ancestry. Specifically, the sample was divided into three groups based on increasing percentages of NA ancestry: 0 to 33% ("low"), 34–66% ("medium"), and 67–100% ("high") NA ancestry. Results from these analyses indicated that, for individuals with the smallest proportion of NA ancestry (and therefore higher

proportion of European ancestry), the PRS for TAG SNPs associated with CPD at $p \leq 0.01$ was positively associated with liability for ND (B_{PRS}=18.9, SE=2.18). In contrast, this association did not hold for individuals with the highest proportion of NA ancestry, and therefore the smallest proportion of European ancestry (B_{PRS}=−1.97, SE=2.98; Fig. 1).

In order to assess whether variants included in the PRS showed differences in LD with surrounding variants as a function of proportion NA ancestry, a paired samples t-test was conducted to compare levels of LD between each SNP (with $p<0.01$) included in the PRS and its correlated SNPs that were identified and removed during the clumping procedure. These LD statistics were calculated separately within the top and bottom thirds of the NA sample. Mean levels of LD were significantly lower for individuals with the highest percent NA ancestry (M=0.041, SD=0.043), compared to individuals with the lowest percent NA ancestry (M=0.055, SD=0.078; $t(65,534)=46.33$, $p<0.001$).

4. DISCUSSION

The current study examined the ability of PRSs computed from a GWAS meta-analysis of CPD in European-ancestry individuals to predict liability for ND in a NA community sample. Risk scores for seven significance thresholds were derived from GWAS test statistics for CPD from the TAG Consortium (TAG, 2010). Each set of scores was used to predict liability for ND in a sample of individuals with varying proportions of NA ancestry. Though no significant main effects were observed, these negative findings were qualified, however, by a significant interaction between the PRS and percent NA ancestry in predicting ND diagnostic status. Specifically, PRSs for CPD were positively associated with liability for ND in higher European-ancestry individuals, and who were therefore more similar in ethnic composition to the discovery sample. In contrast, this relation was not observed in higher NA-ancestry individuals, suggesting that the predictive ability of PRSs derived in a predominantly European-ancestry population may not generalize to individuals with higher proportions of NA ancestry. These findings underscore the need to validate measures of cumulative genetic risk generated from European-ancestry individuals in other racial and ethnic minority populations. If—as observed in the present study—European-ancestry risk scores do not generalize to other ancestral groups, adequately powered primary studies (i.e., GWASs) conducted in these other ancestral groups will be necessary to identify the contributions of individual variants and their patterns of LD with nearby markers before valid PRSs can be constructed for these groups. Given that substance use rates differ across ancestral groups, and health consequences may disproportionately affect specific individuals, these primary studies will also help reduce disparities in what is known about genetic risk factors that contribute to these behaviors in the larger population.

Notably, the lack of predictive power for the PRS among individuals with high NA ancestry appears to have resulted from differences in patterns of LD between NA and European-ancestry individuals. As described, an initial step in creating PRSs involves pruning markers in high LD, so that each locus is tagged by a single variant. If the pattern of LD differs between the discovery sample and the validation sample, a selected variant that tags a specific region in the discovery sample may not effectively tag that region in the validation sample. Consistent with this interpretation, our results demonstrated that in the subset of

participants with the highest degree of NA ancestry (compared to the subset with the highest degree of European ancestry), SNPs included in the PRS showed lower LD with the corresponding markers they were shown to tag during the pruning procedure. This reduced LD was likely responsible for the lack of association between the PRS (for SNPs with $p<0.01$) and liability for ND in the highest percent NA-ancestry individuals. Thus, the present study lends support to arguments suggesting critical information is lost when a pruning approach is used to create PRSs (Vilhjálmsson et al., 2015; Wray et al., 2014).

Alternative methods have recently been developed in order to address this limitation. These include MultiPRS, which improves prediction accuracy by conducting LD pruning separately in discovery samples of different ancestral groups and then creating a PRS in the target sample that is the linear combination of these ancestry-specific PRSs conditional on the individual's own ancestral background (Márquez-Luna et al., 2016). A second method, LDpred, allows for correlated loci in the discovery sample and improves prediction accuracy in two ways: it models the underlying genetic architecture and estimates causal effects sizes for each variant using Bayesian priors, and uses a reference panel to account for LD between associated markers (Vilhjálmsson et al., 2015). The applicability of LDpred for the current study was limited by the small size and extent of admixture present in the target sample, but simulations have demonstrated it represents an improvement over pruning approaches (Vilhjálmsson et al., 2015). The present report provides an empirical demonstration of why it is critical to consider the impact of differences in LD between discovery and target datasets.

A number of limitations of the current study warrant consideration. First, our target sample size was fairly small, and although we were able to represent a wide range of NA ancestry proportions, the relatively small number of individuals may have limited power to detect effects across all p-value thresholds. Nonetheless, the PRS explained a proportion of variance in liability for ND comparable to other smoking phenotypes previously reported using larger samples (Meyers et al., 2013; Vink et al., 2014), and a power analysis indicated the current study was adequately powered (0.82) to detect comparable effect sizes (i.e., $R^2=0.03$). The final model, which included the interaction of the PRS for SNPs with $p<0.01$ and percent NA ancestry, explained 2.2% of the variance in liability for ND. A post-hoc power analysis indicated that with a sample size of $N=288$, there was still reasonable power (0.72) to detect an effect size equivalent to that of the current study ($R^2=0.022$). Second, the present report relied on ND diagnosis rather than CPD to evaluate the predictive power of the PRS. It is possible that the results may have differed if we were able to use the same CPD phenotype was used to generate the PRS. Nonetheless, previous studies have demonstrated a strong genetic correlation between ND and CPD (Vink et al., 2009), suggesting the results would likely be similar across phenotypes.

Despite these limitations, the present study is the first study to apply PRSs derived from SNPs in a predominantly European-ancestry cohort to a sample composed of individuals with varying proportions of NA ancestry. The results suggested that the PRSs might not hold as much predictive accuracy in target samples whose patterns of LD differ from the discovery cohort. This has important implications for future studies that intend to use cumulative measures of risk to predict phenotypes, given that individuals of European ancestry are overrepresented in most large-scale GWASs that typically serve as SNP

discovery datasets. Future investigations might employ newer methods that are able to incorporate more accurate measures of LD from the target population, (MultiPRS; Márquez-Luna et al., 2016), or data from all markers rather than relying on LD-based pruning, (LDpred; Vilhjálmsson et al., 2015), and evaluate how these methods perform across different ancestral populations that differ in their pattern of linkage disequilibrium.

Acknowledgments

The authors thank Helena Furberg and the Tobacco and Genetics (TAG) Consortium for access to their summary data and Derek Wills for his assistance in data management and analysis.

Compliance with Ethical Standards

All procedures involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Data collection and assessment procedures were approved by the Institutional Review Board at The Scripps Research Institute and were also approved by a tribal group overseeing health issues for the communities where recruitment took place. Notably, human subject permissions and the wishes of the participating tribes do not allow study data to be entered into public databases. Informed consent was obtained from all individual participants included in the study.

Role of Funding Source

This research was supported by grants from the National Institute of Drug Abuse (NIDA; DA030976) to IRG, CLE, and KCW and from the National Institute of Alcohol Abuse and Alcoholism (NIAAA) and the National Center on Minority Health and Health Disparities (NCMHD) (5R37 AA010201) to CLE. The funding institutes had no role in study design, in the collection, analysis and interpretation of the data, in the writing of the report, or in the decision to submit the paper for publication.

References

Agrawal A, Lynskey MT. Are there genetic influences on addiction: evidence from family, adoption and twin studies. Addiction. 2008; 103:1069–1081. [PubMed: 18494843]

Agrawal A, Verweij KJH, Gillespie NA, Heath AC, Lessov-Schlaggar CN, Martin NG, Nelson EC, Slutske WS, Whitfield JB, Lynskey MT. The genetics of addiction - a translational perspective. Transl. Psychiatry. 2012; 2:e140. [PubMed: 22806211]

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009; 19:1655–1664. [PubMed: 19648217]

Bizon C, Spiegel M, Chasse SA, Gizer IR, Li Y, Malc EP, Mieczkowski PA, Sailsbery JK, Wang X, Ehlers CL, Wilhelmsen KC. Variant calling in low-coverage whole genome sequencing of a Native American population sample. BMC Genomics. 2014; 15:85. [PubMed: 24479562]

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, Reich T, Schmidt I, Schuckit MA. New semi-structured psychiatric interview for use in genetic linkage studies-report on reliability of SSAGA. J. Stud. Alcohol. 1994; 55:1–10.

Centers for Disease Control and Prevention. Current cigarette smoking among adults—United States, 2005–2014. MMWR. Nov 15.2015 64:1233–1240. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6444a2.htm?s_cid=mm6444a2_w. [PubMed: 26562061]

Chen, L-S.; Saccone, NL.; Culverhouse, RC.; Bracci, PM.; Chen, C-H.; Dueker, N.; Han, Y.; Huang, H.; Jin, G.; Kohno, T.; Ma, JZ.; Przybeck, TR.; Sanders, AR.; Smith, JA.; Sung, YJ.; Wenzlaff, AS.; Wu, C.; Yoon, D.; Chen, Y-T.; Cheng, Y-C.; Cho, YS.; David, SP.; Duan, J.; Eaton, CB.; Furberg, H.; Goate, AM.; Gu, D.; Hansen, HM.; Hartz, S.; Hu, Z.; Kim, YJ.; Kittner, SJ.; Levinson, DF.; Mosley, TH.; Payne, TJ.; Rao, DC.; Rice, JP.; Rice, TK.; Schwantes-An, T-H.; Shete, SS.; Shi, J.; Spitz, MR.; Sun, YV.; Tsai, F-J.; Wang, JC.; Wrensch, MR.; Xian, H.; Gejman, PV.; He, J.; Hunt, SC.; Kardia, SL.; Li, MD.; Lin, D.; Mitchell, BD.; Park, T.; Schwartz, AG.; Shen, H.; Wiencke, JK.; Wu, J-Y.; Yokota, J.; Amos, CI.; Bierut, LJ. Genet. Epidemiol. 36: 2012. Smoking and genetic risk variation across populations of European, Asian, and African American ancestry - a meta-analysis of chromosome 15q25; p. 340-351.

David SP, Hamidovic A, Chen GK, Bergen AW, Wessel J, Kasberger JL, Brown WM, Petruzella S, Thacker EL, Kim Y, Nalls MA, Tranah GJ, Sung YJ, Ambrosone CB, Arnett D, Bandera EV, Becker DM, Becker L, Berndt SI, Bernstein L, Blot WJ, Broeckel U, Buxbaum SG, Caporaso N, Casey G, Chanock SJ, Deming SL, Diver WR, Eaton CB, Evans DS, Evans MK, Fornage M, Franceschini N, Harris TB, Henderson BE, Hernandez DG, Hitsman B, Hu JJ, Hunt SC, Ingles SA, John EM, Kittles R, Kolb S, Kolonel LN, Le Marchand L, Liu Y, Lohman KK, McKnight B, Millikan RC, Murphy A, Neslund-Dudas C, Nyante S, Press M, Psaty BM, Rao DC, Redline S, Rodriguez-Gil JL, Rybicki BA, Signorello LB, Singleton AB, Smoller J, Snively B, Spring B, Stanford JL, Strom SS, Swan GE, Taylor KD, Thun MJ, Wilson AF, Witte JS, Yamamura Y, Yanek LR, Yu K, Zheng W, Ziegler RG, Zonderman AB, Jorgenson E, Haiman CA, Furberg H. Genome-wide meta-analyses of smoking behaviors in African Americans. Transl. Psychiatry. 2012; 2:e119–e118. [PubMed: 22832964]

de Moor MHM, van den Berg SM, Verweij KJH, Krueger RF, Luciano M, Arias Vasquez A, Matteson LK, Derringer J, Esko T, Amin N, Gordon SD, Hansell NK, Hart AB, Seppälä I, Huffman JE, Konte B, Lahti J, Lee M, Miller M, Nutile T, Tanaka T, Teumer A, Viktorin A, Wedenoja J, Abecasis G, Adkins DE, Agrawal A, Allik J, Appel K, Bigdeli TB, Busonero F, Campbell H, Costa PT, Davey Smith G, Davies G, de Wit H, Ding J, Engelhardt BE, Eriksson JG, Fedko IO, Ferrucci L, Franke B, Giegling I, Grucza R, Hartmann AM, Heath AC, Heinonen K, Henders AK, Homuth G, Hottenga JJ, Iacono WG, Janzing J, Jokela M, Karlsson R, Kemp JP, Kirkpatrick MG, Latvala A, Lehtimäki T, Liewald DC, Madden PA, Magri C, Magnusson PKE, Marten J, Maschio A, Medland SE, Mihailov E, Milaneschi Y, Montgomery GW, Nauck M, Ouwens KG, Palotie A, Pettersson E, Polasek O, Qian Y, Pulkki-Råback L, Raitakari OT, Realo A, Rose RJ, Ruggiero D, Schmidt CO, Slutske WS, Sorice R, Starr JM, St Pourcain B, Sutin AR, Timpson NJ, Trochet H, Vermeulen S, Vuoksimaa E, Widén E, Wouda J, Wright MJ, Zgaga L, Porteous D, Minelli A, Palmer AA, Rujescu D, Ciullo M, Hayward C, Rudan I, Metspalu A, Kaprio J, Deary IJ, Räikkönen K, Wilson JF, Keltikangas-Järvinen L, Bierut LJ, Hettema JM, Grabe HJ, van Duijn CM, Evans DM, Schlessinger D, Pedersen NL, Terracciano A, McGue M, Penninx BWJH, Martin NG, Boomsma DI. Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. JAMA Psychiatry. 2015; 72:642–622. [PubMed: 25993607]

de Zeeuw EL, van Beijsterveldt CEM, Glasner TJ, Bartels M, Ehli EA, Davies GE, Hudziak JJ, Genetic Associatio. Rietveld CA, Groen-Blokhuis MM, Hottenga JJ, de Geus EJC, Boomsma DI. Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. Am. J. Med. Genet. 2014; 165:510–520. [PubMed: 25044548]

Domingue BW, Belsky DW, Harris KM, Smolen A, McQueen MB, Boardman JD. Polygenic risk predicts obesity in both white and black young adults. PLoS One. 2014; 9:e101596–e101598. [PubMed: 24992585]

Donny EC, Dierker LC. The absence of DSM-IV nicotine dependence in moderate-to-heavy daily smokers. Drug Alcohol Depend. 2007; 89:93–96. [PubMed: 17276627]

Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013; 9:e1003348–e10033417. [PubMed: 23555274]

Ehlers CL, Gizer IR. Evidence for a genetic component for substance dependence in Native Americans. Am. J. Psychiatry. 2013; 170:154–164. [PubMed: 23377636]

Ehlers CL, Wilhelmsen KC. Genomic screen for loci associated with tobacco usage in Mission Indians. BMC Med. Genet. 2006; 7:9. [PubMed: 16472381]

Hamshere ML, Langley K, Martin J, Agha SS, Stergiakouli E, Anney RJL, Buitelaar J, Faraone SV, Lesch K-P, Neale BM, Franke B, Sonuga-Barke E, Asherson P, Merwood A, Kuntsi J, Medland SE, Ripke S, Steinhausen H-C, Freitag C, Reif A, Renner TJ, Romanos M, Romanos J, Warnke A, Meyer J, Palmason H, Vasquez AA, Lambregts-Rommelse N, Roeyers H, Biederman J, Doyle AE, Hakonarson HH, Rothenberger A, Banaschewski T, Oades RD, McGough JJ, Kent L, Williams N, Owen MJ, Holmans P, O'Donovan MC, Thapar A. High loading of polygenic risk for ADHD in children with comorbid aggression. Am. J. Psychiatry. 2013a; 170:909–916. [PubMed: 23599091]

Hamshere ML, Stergiakouli E, Langley K, Martin J, Holmans P, Kent L, Owen MJ, Gill M, Thapar A, O'Donovan M, Craddock N. Shared polygenic contribution between childhood attention-deficit hyperactivity disorder and adult schizophrenia. Br. J. Psychiatry. 2013b; 203:107–111. [PubMed: 23703318]

Hoffman GE, Mezey JG, Schadt EE. lrgpr: interactive linear mixed model analysis of genome-wide association studies with composite hypothesis testing and regression diagnostics in R. Bioinformatics. 2014; 30:3134–3135. [PubMed: 25035399]

Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011; 27:2987–2993. [PubMed: 21903627]

Libiger O, Schork NJ. A method for inferring an individual's genetic ancestry and degree of admixture associated with six major continental populations. Front. Genet. 2012; 3:322. [PubMed: 23335941]

Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, Berrettini W, Knouff CW, Yuan X, Waeber G, Vollenweider P, Preisig M, Wareham NJ, Zhao JH, Loos RJF, Barroso I, Khaw K-T, Grundy S, Barter P, Mahley R, Kesaniemi A, McPherson R, Vincent JB, Strauss J, Kennedy JL, Farmer A, McGuffin P, Day R, Matthews K, Bakke P, Gulsvik A, Lucae S, Ising M, Brueckl T, Horstmann S, Wichmann H-E, Rawal R, Dahmen N, Lamina C, Polasek O, Zgaga L, Huffman J, Campbell S, Kooner J, Chambers JC, Burnett MS, Devaney JM, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Epstein S, Wilson JF, Wild SH, Campbell H, Vitart V, Reilly MP, Li M, Qu L, Wilensky R, Matthai W, Hakonarson HH, Rader DJ, Franke A, Wittig M, Schäfer A, Uda M, Terracciano A, Xiao X, Busonero F, Scheet P, Schlessinger D, Clair DS, Rujescu D, Abecasis G, Grabe HJ, Teumer A, Völzke H, Petersmann A, John U, Rudan I, Hayward C, Wright AF, Kolcic I, Wright BJ, Thompson JR, Balmforth AJ, Hall AS, Samani NJ, Anderson CA, Ahmad T, Mathew CG, Parkes M, Satsangi J, Caulfield M, Munroe PB, Farrall M, Dominiczak A, Worthington J, Thomson W, Eyre S, Barton A, Mooser V, Francks C, Marchini J. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat. Genet. 2010; 42:436–440. [PubMed: 20418889]

Mackillop J, Obasi EM, Amlung MT, McGeary JE, Knopik VS. The role of genetics in nicotine dependence: Mapping the pathways from genome to syndrome. Curr. Cardiovasc. Risk Rep. 2010; 4:446–453. [PubMed: 21686075]

Márquez-Luna C, Price AL. The SIGMA Type 2 Diabetes Consortium. Multi-ethnic polygenic risk scores improve risk prediction in diverse populations. 2016 doi:http://dx.doi.org/10.1101/051458.

Meyers JL, Cerdá M, Galea S, Keyes KM, Aiello AE, Uddin M, Wildman DE, Koenen KC. Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit neighborhood health study. Transl. Psychiatry. 2013; 3:e290–e299. [PubMed: 23942621]

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 2006; 38:904–909. [PubMed: 16862161]

Rice JP, Hartz SM, Agrawal A, Almasy L, Bennett S, Breslau N, Bucholz KK, Doheny KF, Edenberg HJ, Goate AM, Hesselbrock V, Howells WB, Johnson EO, Kramer J, Krueger RF, Kuperman S, Laurie C, Manolio TA, Neuman RJ, Nurnberger JI, Porjesz B, Pugh E, Ramos EM, Saccone NL, Saccone S, Schuckit M, Bierut LJ. the GENEVA Consortium. CHRNB3 is more strongly associated with FTCD-based nicotine dependence than cigarettes per day: phenotype definition changes genome-wide association studies results. Addiction. 2012; 107:2019–2028. [PubMed: 22524403]

Ripke, S.; Sanders, AR.; Kendler, KS.; Levinson, DF.; Sklar, P.; Holmans, PA.; Lin, D-Y.; Duan, J.; Ophoff, RA.; Andreassen, OA.; Scolnick, E.; Cichon, S.; St Clair, D.; Corvin, A.; Gurling, H.; Werge, T.; Rujescu, D.; Blackwood, DHR.; Pato, CN.; Malhotra, AK.; Purcell, S.; Dudbridge, F.; Neale, BM.; Rossin, L.; Visscher, PM.; Posthuma, D.; Ruderfer, DM.; Fanous, A.; Stefansson, H.; Steinberg, S.; Mowry, BJ.; Golimbet, V.; De Hert, M.; Jönsson, EG.; Bitter, I.; Pietiläinen, OPH.; Collier, DA.; Tosato, S.; Agartz, I.; Albus, M.; Alexander, M.; Amdur, RL.; Amin, F.; Bass, N.; Bergen, SE.; Black, DW.; Borglum, AD.; Brown, MA.; Bruggeman, R.; Buccola, NG.; Byerley, WF.; Cahn, W.; Cantor, RM.; Carr, VJ.; Catts, SV.; Choudhury, K.; Cloninger, CR.; Cormican, P.; Craddock, N.; Danoy, PA.; Datta, S.; De Haan, L.; Demontis, D.; Dikeos, D.; Djurovic, S.; Donnelly, P.; Donohoe, G.; Duong, L.; Dwyer, S.; Fink-Jensen, A.; Freedman, R.; Freimer, NB.; Friedl, M.; Georgieva, L.; Giegling, I.; Gill, M.; Glenthøj, B.; Godard, S.; Hamshere, M.; Hansen, M.; Hansen, T.; Hartmann, AM.; Henskens, FA.; Hougaard, DM.; Hultman, CM.; Ingason, A.; Jablensky, AV.; Jakobsen, KD.; Jay, M.; Jürgens, G.; Kahn, RS.; Keller, MC.; Kenis, G.; Kenny, E.; Kim, Y.; Kirov, GK.; Konnerth, H.; Konte, B.; Krabbendam, L.; Krasucki, R.; Lasseter, VK.;

Laurent, C.; Lawrence, J.; Lencz, T.; Lerer, FB.; Liang, K-Y.; Lichtenstein, P.; Lieberman, JA.; Linszen, DH.; Lönnqvist, J.; Loughland, CM.; Maclean, AW.; Maher, BS.; Maier, W.; Mallet, J.; Malloy, P.; Mattheisen, M.; Mattingsdal, M.; McGhee, KA.; McGrath, JJ.; McIntosh, A.; McLean, DE.; McQuillin, A.; Melle, I.; Michie, PT.; Milanova, V.; Morris, DW.; Mors, O.; Mortensen, PB.; Moskvina, V.; Muglia, P.; Myin-Germeys, I.; Nertney, DA.; Nestadt, G.; Nielsen, J.; Nikolov, I.; Nordentoft, M.; Norton, N.; Nöthen, MM.; O'Dushlaine, CT.; Olincy, A.; Olsen, L.; O'Neill, FA.; Ørntoft, TF.; Owen, MJ.; Pantelis, C.; Papadimitriou, G.; Pato, MT.; Peltonen, L.; Petursson, H.; Pickard, B.; Pimm, J.; Pulver, AE.; Puri, V.; Quested, D.; Quinn, EM.; Rasmussen, HB.; Réthelyi, JM.; Ribble, R.; Rietschel, M.; Riley, BP.; Ruggeri, M.; Schall, U.; Schulze, TG.; Schwab, SG.; Scott, RJ.; Shi, J.; Sigurdsson, E.; Silverman, JM.; Spencer, CCA.; Stefansson, K.; Strange, A.; Strengman, E.; Stroup, TS.; Suvisaari, J.; Terenius, L.; Thirumalai, S.; Thygesen, JH.; Timm, S.; Toncheva, D.; van den Oord, E.; van Os, J.; van Winkel, R.; Veldink, J.; Walsh, D.; Wang, AG.; Wiersma, D.; Wildenauer, DB.; Williams, HJ.; Williams, NM.; Wormley, B.; Zammit, S.; Sullivan, PF.; O'Donovan, MC.; Daly, MJ.; Gejman, PV. Nat. Genet. 43: 2011. Genome-wide association study identifies five new schizophrenia loci; p. 969-976.

Ruderfer DM, Fanous AH, Ripke S, McQuillin A, Amdur RL, Gejman PV, O'Donovan MC, Andreassen OA, Djurovic S, Hultman CM, Kelsoe JR, Jamain S, Landén M, Leboyer M, Nimgaonkar V, Nurnberger J, Smoller JW, Craddock N, Corvin A, Sullivan PF, Holmans P, Sklar P, Kendler KS. Schizophrenia Working Group of the Psychiatric Genetics Consortium, Bipolar Disorder Working Group of the Psychiatric Genetics Consortium, Cross-Disorder Working Group of the Psychiatric Genetics Consortium. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. Mol. Psychiatry. 2013; 19:1017–1024. [PubMed: 24280982]

Saccone NL, Culverhouse RC, Schwantes-An T-H, Cannon DS, Chen X, Cichon S, Giegling I, Han S, Han Y, Keskitalo-Vuokko K, Kong X, Landi MT, Ma JZ, Short SE, Stephens SH, Stevens VL, Sun L, Wang Y, Wenzlaff AS, Aggen SH, Breslau N, Broderick P, Chatterjee N, Chen J, Heath AC, Heliövaara M, Hoft NR, Hunter DJ, Jensen MK, Martin NG, Montgomery GW, Niu T, Payne TJ, Peltonen L, Pergadia ML, Rice JP, Sherva R, Spitz MR, Sun J, Wang JC, Weiss RB, Wheeler W, Witt SH, Yang B-Z, Caporaso N, Ehringer MA, Eisen T, Gapstur SM, Gelernter J, Houlston R, Kaprio J, Kendler KS, Kraft P, Leppert MF, Li MD, Madden PA, Nöthen MM, Pillai SG, Rietschel M, Rujescu D, Schwartz A, Amos CI, Bierut LJ. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. PLoS Genet. 2010; 6:e1001053. [PubMed: 20700436]

Salvatore J, Aliev F, Edwards A, Evans D, Macleod J, Hickman M, Lewis G, Kendler K, Loukola A, Korhonen T, Latvala A, Rose R, Kaprio J, Dick D. Polygenic scores predict alcohol problems in an independent sample and show moderation by the environment. Genes. 2014; 5:330–346. [PubMed: 24727307]

Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, Edenberg HJ, Nurnberger JI, Rietschel M, Blackwood D, Corvin A, Flickinger M, Guan W, Mattingsdal M, McQuillin A, Kwan P, Wienker TF, Daly M, Dudbridge F, Holmans PA, Lin D, Burmeister M, Greenwood TA, Hamshere ML, Muglia P, Smith EN, Zandi PP, Nievergelt CM, McKinney R, Shilling PD, Schork NJ, Bloss CS, Foroud T, Koller DL, Gershon ES, Liu C, Badner JA, Scheftner WA, Lawson WB, Nwulia EA, Hipolito M, Coryell W, Rice JP, Byerley W, McMahon FJ, Schulze TG, Berrettini W, Lohoff FW, Potash JB, Mahon PB, McInnis MG, Zöllner S, Zhang P, Craig DW, Szelinger S, Barrett TB, Breuer R, Meier S, Strohmaier J, Witt SH, Tozzi F, Farmer A, McGuffin P, Strauss J, Xu W, Kennedy JL, Vincent JB, Matthews K, Day R, Ferreira MD, O'Dushlaine C, Perlis R, Raychaudhuri S, Ruderfer D, Hyoun PL, Smoller JW, Li J, Absher D, Thompson RC, Meng FG, Schatzberg AF, Bunney WE, Barchas JD, Jones EG, Watson SJ, Myers RM, Akil H, Boehnke M, Chambert K, Moran J, Scolnick E, Djurovic S, Melle I, Morken G, Gill M, Morris D, Quinn E, Mühleisen TW, Degenhardt FA, Mattheisen M, Schumacher J, Maier W, Steffens M, Propping P, Nöthen MM, Anjorin A, Bass N, Gurling H, Kandaswamy R, Lawrence J, Mcghee K, McIntosh A, McLean AW, Muir WJ, Pickard BS, Breen G, St Clair D, Caesar S, Gordon-Smith K, Jones L, Fraser C, Green EK, Grozeva D, Jones IR, Kirov G, Moskvina V, Nikolov I, O'Donovan MC, Owen MJ, Collier DA, Elkin A, Williamson R, Young AH, Ferrier IN, Stefansson K, Stefansson H, Porgeirsson P, Steinberg S, Gustafsson O, Bergen SE, Nimgaonkar V, hultman C, Landén M, Lichtenstein P, Sullivan P, Schalling M, Osby U, Backlund L, Frisén L, Långström N, Jamain S, Leboyer M, Etain B, Bellivier F, Petursson H, Sigur Sson E, Müller-Mysok B, Lucae S, Schwarz M, Schofield PR, Martin N, Montgomery GW, Lathrop M, Oskarsson H, Bauer M, Wright A,

Mitchell PB, Hautzinger M, Reif A, Kelsoe JR, Purcell SM. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat. Genet. 2011; 43:977–983. [PubMed: 21926972]

The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 490:56–65.

The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460:748–752. [PubMed: 19571811]

Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsäter A, Flex A, Aben KKH, de Vegt F, Mulders PFA, Isla D, Vidal MJ, Asin L, Saez B, Murillo L, Blondal T, Kolbeinsson H, Stefansson JG, Hansdottir I, Runarsdottir V, Pola R, Lindblad B, van Rij AM, Dieplinger B, Haltmayer M, Mayordomo JI, Kiemeney LA, Matthiasson SE, Oskarsson H, Tyrfingsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A, Stefansson K. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008; 452:638–642. [PubMed: 18385739]

Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, Gieger C, Rawal R, Mangino M, Prokopenko I, Mägi R, Keskitalo K, Gudjonsdottir IH, Gretarsdottir S, Stefansson H, Thompson JR, Aulchenko YS, Nelis M, Aben KK, Heijer den M, Dirksen A, Ashraf H, Soranzo N, Valdes AM, Steves C, Uitterlinden AG, Hofman A, Tönjes A, Kovacs P, Hottenga JJ, Willemsen G, Vogelzangs N, Döring A, Dahmen N, Nitz B, Pergadia ML, Saez B, De Diego V, Lezcano V, Garcia-Prats MD, Ripatti S, Perola M, Kettunen J, Hartikainen A-L, Pouta A, Laitinen J, Isohanni M, Huei-Yi S, Allen M, Krestyaninova M, Hall AS, Jones GT, van Rij AM, Mueller T, Dieplinger B, Haltmayer M, Jonsson S, Matthiasson SE, Oskarsson H, Tyrfingsson T, Kiemeney LA, Mayordomo JI, Lindholt JS, Pedersen JH, Franklin WA, Wolf H, Montgomery GW, Heath AC, Martin NG, Madden PA, Giegling I, Rujescu D, Järvelin M-R, Salomaa V, Stumvoll M, Spector TD, Wichmann H-E, Metspalu A, Samani NJ, Penninx BW, Oostra BA, Boomsma DI, Tiemeier H, van Duijn CM, Kaprio J, Gulcher JR, McCarthy MI, Peltonen L, Thorsteinsdottir U, Stefansson K. Sequence variants at CHRNB3–CHRNA6 and CYP2A6 affect smoking behavior. Nat. Genet. 2010; 42:448–453. [PubMed: 20418888]

Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat. Genet. 2010; 42:441–447. [PubMed: 20418890]

U.S. Department of Health and Human Services. Tobacco use among U.S. racial/ethnic minority groups—African Americans, American Indians and Alaska Natives, Asian Americans, and Hispanics: A report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 1998. Available at http://www.cdc.gov/tobacco/data_statistics/sgr/1998/complete_report/pdfs/complete_report.pdf [Accessed 10 October 2015]

U.S. Department of Health and Human Services. How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease: A report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Healt; 2010. Available at http://www.cdc.gov/tobacco/data_statistics/sgr/2010/index.htm [Accessed 7 October 2015]

U.S. Department of Health and Human Services. The health consequences of smoking— 50 years of progress: A report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014. Available at http://www.surgeongeneral.gov/library/reports/50-years-of-progress/full-report.pdf [Accessed 9 October 2015]

Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, Genovese G, Loh P-R, Bhatia G, Do R, Hayeck T, Won H-H, Schizophrenia Working Group of the Psychiatric Genomics Consortium. Kathiresan S, Pato M, Pato C, Tamimi R, Stahl E, Zaitlen N, Pasaniuc B, Belbin G, Kenny EE, Schierup MH, De Jager P, Patsopoulos NA, McCarroll S, Daly M, Purcell S, Chasman

D, Neale B, Goddard M, Visscher PM, Kraft P, Patterson N, Price AL. Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet. 2015; 97:576–592. [PubMed: 26430803]

Vink JM, Hottenga JJ, de Geus EJC, Willemsen G, Neale MC, Furberg H, Boomsma DI. Polygenic risk scores for smoking: predictors for alcohol and cannabis use? Addiction. 2014; 109:1141–1151. [PubMed: 24450588]

Vink JM, Smit AB, de Geus EJC, Sullivan P, Willemsen G, Hottenga JJ, Smit JH, Hoogendijk WJ, Zitman FG, Peltonen L, Kaprio J, Pedersen NL, Magnusson PK, Spector TD, Kyvik KO, Morley KI, Heath AC, Martin NG, Westendorp RGJ, Slagboom PE, Tiemeier H, Hofman A, Uitterlinden AG, Aulchenko YS, Amin N, van Duijn C, Penninx BW, Boomsma DI. Genome-wide association study of smoking initiation and current smoking. Am. J. Hum. Genet. 2009; 84:367–379. [PubMed: 19268276]

Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. Behav. Genet. 2005; 35:397–406. [PubMed: 15971021]

Vrieze SI, McGue M, Iacono WG. The interplay of genes and adolescent development in substance use disorders: Leveraging findings from GWAS meta-analyses to test developmental hypotheses about nicotine consumption. Hum. Genet. 2012; 131:791–801. [PubMed: 22492059]

Wall TL, Carr LG, Ehlers CL. Protective association of genetic variation in alcohol dehydrogenase with alcohol dependence in Native American Mission Indians. Am. J. Psychiatry. 2003; 160:41–46. [PubMed: 12505800]

Ware JJ, van den Bree MBM, Munafo MR. Association of the CHRNA5-A3-B4 gene cluster with heaviness of smoking: a meta-analysis. Nicotine Tob. Res. 2011; 13:1167–1175. [PubMed: 22071378]

Wilhelmsen KC, Ehlers C. Heritability of substance dependence in a Native American population. Psychiatr. Genet. 2005; 15:101–107. [PubMed: 15900224]

Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res. 2007; 17:1520–1528. [PubMed: 17785532]

Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research Review: polygenic methods and their application to psychiatric traits. J. Child Psychol. Psychiatry. 2014; 55:1068–1087. [PubMed: 25132410]

Highlights

- Polygenic risk scores (PRS) represent additive effects of common genetic variants
- PRS for cigarettes per day (CPD) were derived from European ancestry GWAS data
- PRS predicted liability for nicotine dependence (ND) in European ancestry subjects
- PRS did not predict liability for ND in Native American ancestry subjects
- Model linkage disequilibrium when creating PRS with different ancestral populations

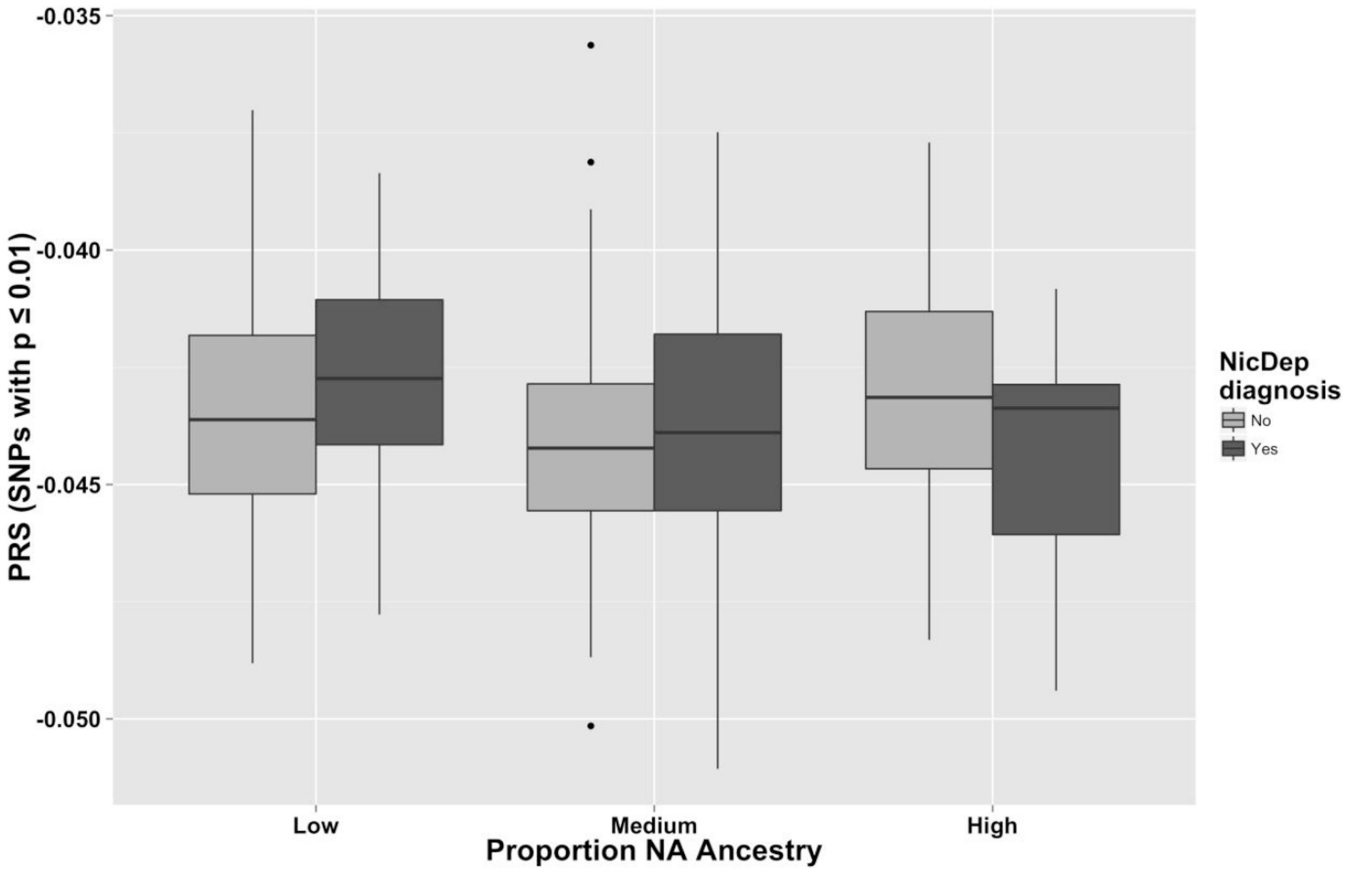

Fig. 1.

Moderating effect of percent Native American ancestry on the relation between polygenic risk score (for SNPs with $p \leq 0.01$) and nicotine dependence diagnosis

Table 1

Main effect of polygenic risk score at each inclusion threshold on liability for nicotine dependence diagnosis.

p-value inclusion threshold	Number of SNPs	B_{PRS}	SE	*t*	*p*	R^2
0.5	240,040	−1.86	9.72	−0.19	0.8479	0.000
0.4	199,384	−2.34	8.55	−0.03	0.9782	0.000
0.3	155,598	−1.84	7.12	−0.26	0.7964	0.000
0.2	109,056	−4.14	5.59	−0.74	0.4588	0.002
0.1	58,591	25.49	3.78	0.68	0.4996	0.002
0.05	31,240	17.53	2.55	0.69	0.4910	0.002
0.01	7,235	11.56	1.17	0.99	0.3208	0.003

B_{PRS}: Regression coefficient of polygenic risk score

Table 2

Linear mixed model analysis of the effects of polygenic risk score and percent Native American ancestry on liability for nicotine dependence diagnosis.

Variable	B	SE	*t*	*p*
Model with main effects of percent Native American ancestry and polygenic risk score (SNPs with $p \leq 0.01$)				
Sex	−0.11	5.61e-02	−1.96	0.0500
Age	0.02	9.09e-03	2.59	0.0096
Age-squared	−2.66e-04	1.12e-04	−2.39	0.0170
European ancestry	0.56	4.44e-01	1.26	0.2083
African ancestry	0.13	6.37e-01	0.20	0.8423
Native American ancestry	0.11	4.48e-01	0.24	0.8085
PRS for SNPs $p \leq 0.01$	11.56	1.17	0.99	0.3208
Model with main effects of percent Native American ancestry and polygenic risk score and their interaction (SNPs with $p \leq 0.01$)				
Sex	−9.83e-02	5.58e-02	−1.76	0.0783
Age	2.34e-02	9.00e-03	2.59	0.0095
Age-squared	−2.64e-04	1.11e-04	−2.39	0.0168
European ancestry	5.07e-01	4.40e-01	1.15	0.2495
African ancestry	−2.04e-02	6.34e-01	−0.03	0.9743
Native American ancestry	**−5.64**	**2.49**	**−2.26**	**0.0237** *
PRS for SNPs $p \leq 0.01$	**7.05**	**2.77**	**2.55**	**0.0109** *
Native American ancestry x PRS ($p \leq 0.01$)	**−1.31**	**5.60**	**−2.34**	**0.0191** *

PRS: polygenic risk score; SNPs: single nucleotide polymorphisms

* $p<0.05$