White Matter Integrity and Nicotine Dependence: Evaluating Vertical and Horizontal Pleiotropy

Zhenyao Ye[1,2†], Chen Mo[1,2†], Song Liu[3†], Kathryn S. Hatch[1], Si Gao[1], Yizhou Ma[1], L. Elliot Hong[1], Paul M. Thompson[4], Neda Jahanshad[4], Ashley Acheson[5], Hugh Garavan[6], Li Shen[7], Thomas E. Nichols[8], Peter Kochunov[1*†], Shuo Chen[1,2*†] and Tianzhou Ma[9*†]

[1] Maryland Psychiatric Research Center, Department of Psychiatry, School of Medicine, University of Maryland, Baltimore, Baltimore, MD, United States, [2] Division of Biostatistics and Bioinformatics, Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, Baltimore, MD, United States, [3] School of Computer Science and Technology, Qilu University of Technology, Shandong Academy of Sciences, Jinan, China, [4] Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States, [5] Department of Psychiatry and Behavioral Sciences, University of Arkansas for Medical Sciences, Little Rock, AR, United States, [6] Department of Psychiatry, The University of Vermont, Burlington, VT, United States, [7] Department of Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, [8] Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom, [9] Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, College Park, MD, United States

Tobacco smoking is an addictive behavior that supports nicotine dependence and is an independent risk factor for cancer and other illnesses. Its neurogenetic mechanisms are not fully understood but may act through alterations in the cerebral white matter (WM). We hypothesized that the vertical pleiotropic pathways, where genetic variants influence a trait that in turn influences another trait, link genetic factors, integrity of cerebral WM, and nicotine addiction. We tested this hypothesis using individual genetic factors, WM integrity measured by fractional anisotropy (FA), and nicotine dependence-related smoking phenotypes, including smoking status (SS) and cigarettes per day (CPDs), in a large epidemiological sample collected by the UK Biobank. We performed a genome-wide association study (GWAS) to identify previously reported loci associated with smoking behavior. Smoking was found to be associated with reduced WM integrity in multiple brain regions. We then evaluated two competing vertical pathways: Genes → WM integrity → Smoking versus Genes → Smoking → WM integrity and a horizontal pleiotropy pathway where genetic factors independently affect both smoking and WM integrity. The causal pathway analysis identified 272 pleiotropic single-nucleotide polymorphisms (SNPs) whose effects on SS were mediated by FA, as well as 22 pleiotropic SNPs whose effects on FA were mediated by CPD. These SNPs were mainly located in important susceptibility genes for smoking-induced diseases NCAM1 and IREB2. Our findings revealed the role of cerebral WM in the maintenance of the complex addiction and provided potential genetic targets for future research in examining how changes in WM integrity contribute to the nicotine effects on the brain.

Keywords: nicotine, white matter integrity, pleiotropy, mediation analysis, causal inference

INTRODUCTION

Tobacco smoking is a complex addictive behavior and is the chief modifiable causal factor for cancer, coronary and cerebrovascular disorders, chronic obstructive pulmonary disorder, hypertension, and many other illnesses (McBride, 1992; Ziedonis et al., 2008). Additive genetic factors explain up to ∼75% of the variance on traits that quantify nicotine dependence and smoking-related behaviors in population studies, thus supporting the hypothesis of a strong genetic susceptibility for nicotine addiction (Kendler et al., 1999; Vink et al., 2005). Genome-wide association studies (GWASs) localized and replicated multiple genetic variants conferring susceptibility to smoking/nicotine addiction including those regulating nicotinic acetylcholine receptors (nAChRs) (True et al., 1999; Erzurumluoglu et al., 2019). However, its neurogenetic mechanisms remain unknown. Neuroimaging studies reported observational (Zhang et al., 2010; Gons et al., 2011; Gray et al., 2020) and direct associations (Kochunov et al., 2013) between smoking and nicotine administration and white matter (WM) integrity, assessed by the fractional anisotropy (FA) of water diffusion in diffusion tensor imaging (DTI) (Kochunov et al., 2007). Cerebral WM is hypothesized to be involved in both positive and negative reinforcement mechanisms of nicotine addiction (Benowitz, 2010). The positive effect of nicotine is associated with cognitive and mood enhancement, which could be driven through the transient elevation in WM integrity following nicotine administration (Perkins, 1999; Kochunov et al., 2013). The nAChRs have been found in the cerebral WM of both human and nonhuman primates in the previous PET ligand and histological binding studies (Fujita et al., 2003; Ding et al., 2004; Pimlott et al., 2004; Hillmer et al., 2011). Their functions have been studied in peripheral (Lang et al., 2003) and central (Zhang et al., 2011a) nervous systems. The negative effect of nicotine in avoiding withdrawal symptoms may also be caused by reduction in WM integrity due to cerebrovascular and neurodegenerative risks associated with heavy chronic smoking (Hudkins et al., 2010; Kim et al., 2010; Gons et al., 2011; Liao et al., 2011; Zhang et al., 2011a).

Vertical and horizontal pleiotropic pathways are chief causal mechanisms explaining the phenomenon of genetic variants affecting multiple traits (Tyler et al., 2009; Paaby and Rockman, 2013; Jordan et al., 2019). Vertical pleiotropy is observed when a trait influenced by genetic factors has, in turn, influenced another trait by acting as a mediator. Horizontal pleiotropy refers to two traits being independently influenced by the same genetic factors. In the first vertical pleiotropic pathway (**Figure 1**: model 1), we assume that genetic factors underpinning susceptibility to nicotine dependence do so through alterations in WM integrity. In the second vertical pleiotropic pathway (**Figure 1**: model 2), we alternatively hypothesize that genetic factors underpin nicotine addiction directly, and the reduction in WM integrity occurs due to a multitude of harmful effects of chronic or heavy smoking. Lastly, we also test horizontal pleiotropy (**Figure 1**: model 0) where genetic factors affect both WM integrity and nicotine addiction independently. The proposed models are complementary to the widely used Mendelian randomization (MR) models that estimate the causal influence between the traits by relaxing the conditional independence assumption in instrumental variable (IV) analysis (Davey Smith and Hemani, 2014; Hemani et al., 2018a,b). We then develop rigorous analytical approaches to determine and validate the best causal model for pleiotropic genetic variants associated with both WM integrity and nicotine addiction.

We analyzed two smoking phenotypes for their biological relevance to nicotine dependence, smoking status (SS) and cigarettes per day (CPD), to test the proposed causal pathways in large-scale epidemiological data from the UK Biobank (UKBB). We used FA to measure WM integrity. Smoking (being current smoker or having higher CPD) was found to be associated with reduced WM integrity in multiple brain regions. Our data showed that the genetic effects on FA and the two smoking phenotypes were not independent, so the horizontal pleiotropy does not hold. Since model 1 and model 2 with a vertical pleiotropic relationship are two mutually exclusive causal pathways, we performed mediation analysis and used Bayes factor to select the optimal model. We identified 272 pleiotropic variants associated with both SS and FA whose effects on SS were mediated by FA (model 1 preferred). On the contrary, 22 pleiotropic variants were found to be associated with both CPD and FA, where CPD acts as a mediator for the genetic effects on FA (model 2 preferred). The identified variants mainly reside in two genes *NCAM1* and *IREB2*. Their relationship to the smoking-induced brain mechanisms will need to be further examined for their functionality in future studies.
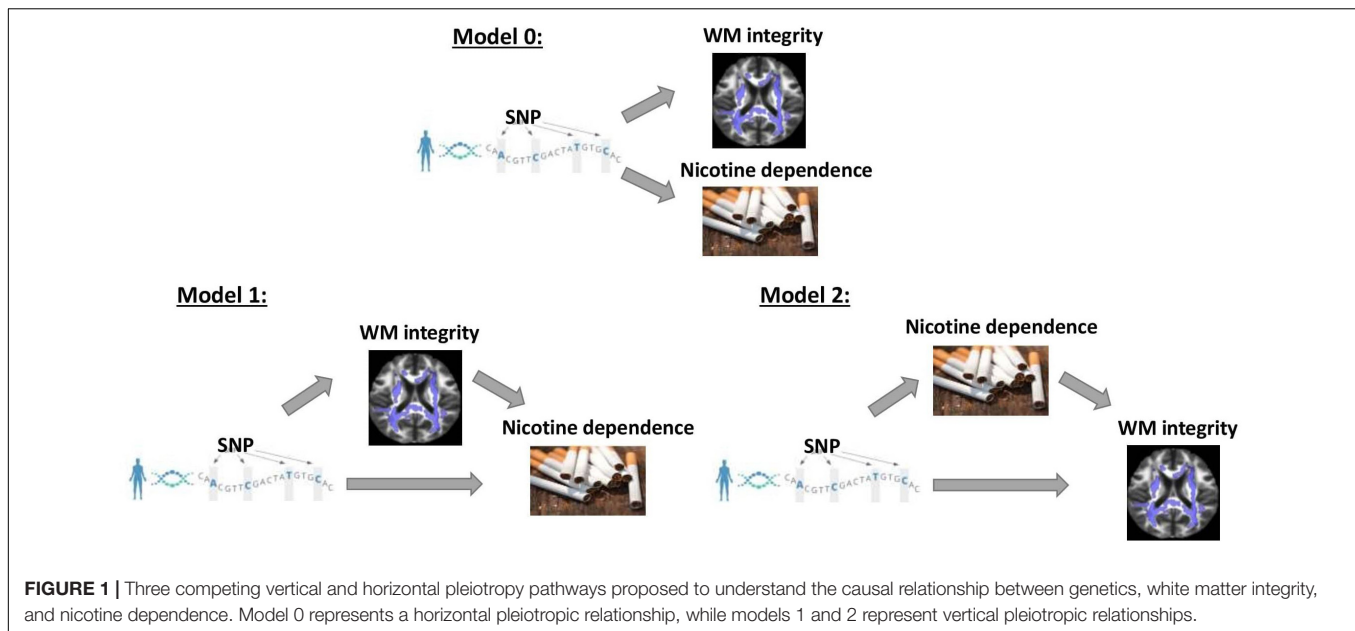
MATERIALS AND METHODS

UK Biobank Cohort

The data used to test our causal pathways are from the UKBB, a large prospective study that recruited ∼500,000 participants aged between 40 and 69 years in 2006–2010 in 22 assessment centers throughout the United Kingdom (Sudlow et al., 2015). UKBB data consist of phenotypic, genotypic, and imaging details about its participants collected from questionnaires, physical measures, multimodal imaging, genome-wide genotyping, and longitudinal follow-up for health-associated outcomes (Sudlow et al., 2015). In our analysis, UKBB data from all sites and all phases are included. UKBB carried out an automated processing pipeline for sample retrieval, data collection, and quality control to convert raw data to reliable processed data for use by all researchers (UK Biobank, 2007; Alfaro-Almagro et al., 2018; Bycroft et al., 2018). We restricted our analysis to only participants with white ethnicity backgrounds (British, Irish, and any other white background) and with both genotype and nicotine dependence phenotype data available. For causal pathway analysis, we further narrowed down to participants who have genotype, nicotine dependence, and WM integrity phenotype data available. The number of participants included at each analytic step is summarized in **Supplementary Figure 1**.

Nicotine Dependence-Related Smoking Phenotypes

Supplementary Table 1 summarizes the number of participants by smoking-related phenotype codes in UKBB. We chose to analyze the following two phenotypes due to their

FIGURE 1 | Three competing vertical and horizontal pleiotropy pathways proposed to understand the causal relationship between genetics, white matter integrity, and nicotine dependence. Model 0 represents a horizontal pleiotropic relationship, while models 1 and 2 represent vertical pleiotropic relationships.

biological relevance to nicotine dependence (Donny et al., 2008; Benowitz, 2010):

(1) Smoking Status (SS; current vs. never smokers). SS was a binary trait describing subjects who were either current smokers or never smokers (past smokers excluded). Current and never smokers were defined using phenotype code 20116 (smoking status) in UKBB.

(2) Cigarettes Per Day (CPD; average number of cigarettes smoked per day by participants who are either current or past smokers). CPD was a quantitative trait that measures the heaviness of smoking among ever smokers (never smokers excluded). It was defined using phenotype codes 2887 (number of cigarettes previously smoked daily), 3456 [number of cigarettes currently smoked daily), and 6183 (number of cigarettes previously smoked daily (current cigar/pipe smokers)] in UKBB. The CPD values of participants who smoked less than one CPD were recoded to 0; and CPD values of those who smoked more than 60 CPD were recoded to 60.

Note that SS and CPD are both related to nicotine dependence but target different smoker categories. SS focuses on the difference between current smokers and never smokers, while CPD focuses on the heaviness of smoking among current or past smokers.

White Matter Integrity Phenotype Measured by Fractional Anisotropy

The UKBB consists of multimodal braining imaging data covering structural, functional, and diffusion imaging (Miller et al., 2016). In this study, we concentrated on the WM FA measure derived from diffusion MRI data, a common measure of WM integrity whose association with smoking addiction behavior has been reported in previous studies

(Gogliettino et al., 2016). The UKBB database provides 40 FA measures from multiple brain regions (full region names listed in **Supplementary Table 2**), including the inferior cerebellar peduncle (ICP), genu of the corpus callosum (GCC), body of the corpus callosum (BCC), splenium of the corpus callosum (SCC), fornix (FX), corticospinal tract (CST) (mean/right/left), anterior limb of the internal capsule (ALIC) (right/left), posterior limb of the internal capsule (PLIC) (right/left), retrolenticular part of the internal capsule (RLIC) (right/left), anterior corona radiata (ACR) (right/left), superior corona radiata (SCR) (right/left), posterior corona radiata (PCR) (right/left), posterior thalamic radiation (PTR) (right/left), sagittal striatum (SS) (right/left), external capsule (EX) (right/left), cingulum cingulate gyrus (CGC) (right/left), cingulum hippocampus (CHG) (right/left), fornix cres+stria terminalis (FXST) (right/left), superior longitudinal fasciculus (SLF) (right/left), superior fronto-occipital fasciculus (SFO) (right/left), uncinate fasciculus (UN) (right/left), and tapetum (TAP) (right/left). The data were preprocessed using pipeline similar to that developed by Enhancing Neuro Imaging Genetics Meta Analysis (ENIGMA) consortium (Thompson et al., 2014). Before conducting the causal pathway analysis for each FA, we further performed univariate association analysis by regressing FA measures on the two smoking phenotypes and kept only those FA measures with lower values among current smokers or negatively associated with CPD ($\hat{\beta} < 0$, $p < 0.05$), motivated by the findings from previous studies (Savjani et al., 2014; Umene-Nakano et al., 2014; Gray et al., 2020).

Genotype Data and Genome-Wide Association Study (GWAS)

The genotype data of UKBB cohort came from two platforms, Affymetrix UK BiLEVE Axiom and UKBB Axiom® arrays, which captured over 90 million single-nucleotide variants (SNVs) of

~500,000 subjects (Bycroft et al., 2017). We first removed variants with minor allele frequency (MAF) below 0.01, Hardy–Weinberg equilibrium p-value below 0.001, and missing genotype rate at 5% and removed individuals with more than 2% missing genotypes. We conducted principal component analysis (PCA) method to adjust for population stratification and chose the top 10 principle components (PCs) as recommended by PLINK (version 2.0)[1] (Chang et al., 2015) and in previous studies (Feng et al., 2009; Kang et al., 2009; Price et al., 2010; Warren et al., 2017). We then performed GWAS separately on the two smoking phenotypes SS and CPD adjusting for age, gender, body mass index, genotyping chip type, and the acquired top 10 PCs using PLINK (version 1.9[2]) (Chang et al., 2015). The most significantly associated loci (i.e., $p < 5 \times 10^{-8}$) were used for the causal pathway analysis. Considering the strong linkage disequilibrium (LD) of neighboring single-nucleotide polymorphisms (SNPs), we also included SNPs in the genomic regions $\pm$ 250 kb around the peak boundaries (**Supplementary Figures 2–7**). GWAS was not performed on FA measures, which is underpowered due to the relatively small number of participants having neuroimaging data available. In the causal pathway analysis, we further selected pleiotropic SNPs that are associated with both smoking phenotypes and FA measures.

Causal Pathway Discovery

Denote by G the genotype, M the FA measures, Y the smoking phenotypes (SS or CPD), and Z the potential covariates. In this study, we included age and gender as covariates to be adjusted in the model. Given the directed graph structures in **Figure 1**, we can represent the three competing models by factorizing their joint distributions:

$$Model 0 : Pr(M, Y|G, Z) = Pr(M|(G, Z))Pr(Y|(G, Z))$$

$$Model 1 : Pr(M, Y|G, Z) = Pr(M|(G, Z))Pr(Y|(M, G, Z))$$

$$Model 2 : Pr(M, Y|G, Z) = Pr(Y|(G, Z))Pr(M|(Y, G, Z))$$

Model 0 assumes genetics to be a common cause of both FA and smoking phenotypes independently ("SNP → Smoking," "SNP → FA" and FA ⊥ Smoking given SNP and confounders) and represents a horizontal pleiotropic relationship. Model 1 and model 2 are two alternative mediation models that represent a vertical pleiotropic relationship. In model 1 ("SNP → FA → Smoking"), FA measures are regarded as the mediators that mediate the effect of SNPs on smoking. In contrast, model 2 ("SNP → Smoking → FA") considers the long-term effect of chronic smoking on the brain structure and regards smoking as the mediator that mediates the effect of SNPs on FA.

We performed automatic causal pathway discovery analysis (Heckerman et al., 1999; Spirtes and Zhang, 2016; Glymour et al., 2019) to identify the optimal pleiotropic pathway. Our analyses started by identifying the potentially pleiotropic variants of FA and smoking for each FA measure separately. We then

evaluated the association between each FA measure and smoking given the SNP effects to distinguish horizontal pleiotropy from vertical pleiotropy. For variants with a vertical pleiotropic relationship, we further conducted causal mediation analysis to choose the best mediation model that explains the relationship between SNP, FA, and smoking. Below, we describe the analytical steps in details.

Step 1: Identification of Pleiotropic Variants

Suppose we start with a set of g_0 SNPs gained by GWAS for n subjects. Let G_{ij} denote the genotype of the ith subject in the jth SNP ($1 \leq i \leq n, j \in g_0$), M_{il} denote the lth continuous FA measure ($1 \leq l \leq 40$) of ith subject, Y_i denote the SS/CPD value, and Z_i denote the covariates of the ith subject. We assume an additive genetic model, and let $G_{ij} = 0$, 1, or, 2 represent the number of copies of minor alleles. In the first step, we look for SNPs that are associated with both FA measures and SS/CPD (i.e., potentially pleiotropic variants). Notably, this step is also a necessary condition to establish mediation for both model 1 and model 2, where the mediator and the outcome are simply switched in the two models (Judd and Kenny, 1981; James and Brett, 1984; Baron and Kenny, 1986). First, we fit a linear regression model on each SNP for each FA measure separately adjusting for the covariates Z to look for SNP–FA association. Then we fit a logistic regression or a linear regression model on each SNP for SS or CPD, adjusting for the covariates Z to look for SNP–smoking association:

(1a) Regress M on G and Z:

$$M_{il} = \alpha_1 + \beta_{1j}G_{ij} + \gamma_1 Z_i + \varepsilon_{1i}, \ \varepsilon_{1i} \sim N\left(0, \sigma_1^2\right),$$
$$1 \leq i \leq n, \ 1 \leq l \leq 40, \ j \in g_0$$

(1b) Regress Y on G and Z:

$$Y \text{ is binary (SS) : logit}(P(Y_i = 1)) = \alpha_2 + \beta_{2j}G_{ij} + \gamma_2 Z_i,$$
$$1 \leq i \leq n, \ j \in g_0$$

$$Y \text{ is continuous (CPD) :} Y_i = \alpha_2 + \beta_{2j}G_{ij} + \gamma_2 Z_i + \varepsilon_{2i}, \ \varepsilon_{2i}$$
$$\sim N\left(0, \sigma_2^2\right), \ 1 \leq i \leq n, \ j \in g_0$$

where α_1 and α_2 are the intercepts, γ_1 and γ_2 are the effects of covariates, and β_{1j} and β_{2j} correspond to the genetic effect of the jth SNP on M and Y (subscript l was omitted for α_1 and β_{1j} for simplicity). The cutoff for statistical significance is chosen to control for the overall false discovery rate (FDR < 0.15) in identifying the potentially pleiotropic variants in the shared subset that meets both SNP–FA and SNP–smoking association criteria (see **Supplementary Methods**). We then used the Functional Annotation of Variants—Online Resource (FAVOR) (Li et al., 2020) to annotate the identified pleiotropic variants for more biological insights.

Step 2: Distinguish Horizontal From Vertical Pleiotropy

Model 0 assumes a horizontal pleiotropic relationship, while models 1 and 2 assume a vertical pleiotropic relationship. The

main feature of horizontal pleiotropy is that the two traits (smoking and FA) are independent given the SNP effect. In this step, we conducted association analysis between FA and SS/CPD conditioning on the SNP. If the conditional independence holds (conditional independence test $p > 0.05$), the variants demonstrate horizontal pleiotropy; otherwise, they demonstrate vertical pleiotropy.

Step 3: Selection of the Best Mediation Model for Vertical Pleiotropy

Model 1 and model 2 are two mutually exclusive mediation models under the vertical pleiotropy assumption. Mediation analysis investigates how a third variable affects the relation between two other variables and is a useful tool in discovering the hidden mechanism in many biological fields (MacKinnon et al., 2007). We conducted exploratory mediation analysis and selected the best mediation model for each pleiotropic SNP using Bayes factor criteria (Kass and Raftery, 1995; Denison et al., 2002; Bernardo and Smith, 2009). We validated the selected model by checking the major causality assumptions and testing and categorizing the mediation effects.

Suppose that the set of SNPs that meet the criteria of vertical pleiotropy from step 1 and 2 is g_1, to determine the best mediation model from the two candidates, and we regress the outcome on both exposure variable and mediator for model 1 (outcome = SS/CPD) and model 2 (outcome = FA), respectively, adjusting for covariates Z:

(2a) Regress Y on G, M, and Z:

$$Y \text{ is binary (SS)} : \text{logit}(P(Y_i = 1)) = \alpha_3 + \beta_{31j}G_{ij} + \beta_{32}M_{il}$$
$$+\gamma_3 Z_i,\ 1 \le i \le n,\ 1 \le l \le 40,\ j \in g_1$$

$$Y \text{ is continuous (CPD)} : Y_i = \alpha_3 + \beta_{31j}G_{ij} + \beta_{32}M_{il} + \gamma_3 Z_i$$
$$+\varepsilon_{3i},\ \varepsilon_{3i} \sim N\left(0, \sigma_3^2\right),\ 1 \le i \le n, 1 \le l \le 40,\ j \in g_1$$

(2b) Regress M on G, Y, and Z:

$$M_{il} = \alpha_4 + \beta_{41j}G_{ij} + \beta_{42}Y_i + \gamma_4 Z_i + \varepsilon_{4i},\ \varepsilon_{4i} \sim N\left(0, \sigma_4^2\right),$$
$$1 \le i \le n, 1 \le l \le 40,\ j \in g_1$$

where α_3 and α_4 are the intercepts; γ_3 and γ_4 are the effects of covariates; β_{31j} and β_{41j} represent the direct effects of SNPs on outcomes Y and M in models 1 and 2, respectively; and $\beta_{1j}\beta_{32}$ and $\beta_{2j}\beta_{42}$ represent the indirect effects of SNPs on outcome via the mediators M or Y in models 1 and 2, respectively.

To select the mediation model that best explains the causal relationship for each SNP j, we propose to use the penalized likelihood Bayesian information criterion (BIC) score as a model selection criterion (Schwarz, 1978). By definition, BIC $= -2\log(\hat{L}) + p\log(n)$, where $\hat{L}$ denotes the maximized value of the likelihood, p is the number of parameters, and n is the sample size. Since the two mediation models have exactly the same number of parameters, we are directly comparing the maximum likelihoods of the two models. The maximum likelihoods of model 1 and model 2 can be derived from their joint distribution combining

step 1 and step 2 (model 1: 1a+2a; model 2: 1b+2b), with the parameters evaluated at maximum likelihood estimation (MLE):

$$\text{Model1} : \hat{L}(M1) = L(\widehat{\alpha_1}, \widehat{\beta_{1j}},\ \widehat{\gamma_1}, \widehat{\sigma_1^2}|M_l, G_j, Z)$$
$$L(\widehat{\alpha_3}, \widehat{\beta_{31j}}, \widehat{\beta_{32j}},\ \widehat{\gamma_3}, \widehat{\sigma_3^2}|Y,\ G_j, M_l, Z)$$

$$\text{Model2} : \hat{L}(M2) = L(\widehat{\alpha_2}, \widehat{\beta_{2j}},\ \widehat{\gamma_2}, \widehat{\sigma_2^2}|Y, G_j, Z)$$
$$L(\widehat{\alpha_4}, \widehat{\beta_{41j}}, \widehat{\beta_{42j}}\ \widehat{\gamma_4}, \widehat{\sigma_4^2}|M_l,\ G_j, Y, Z)$$
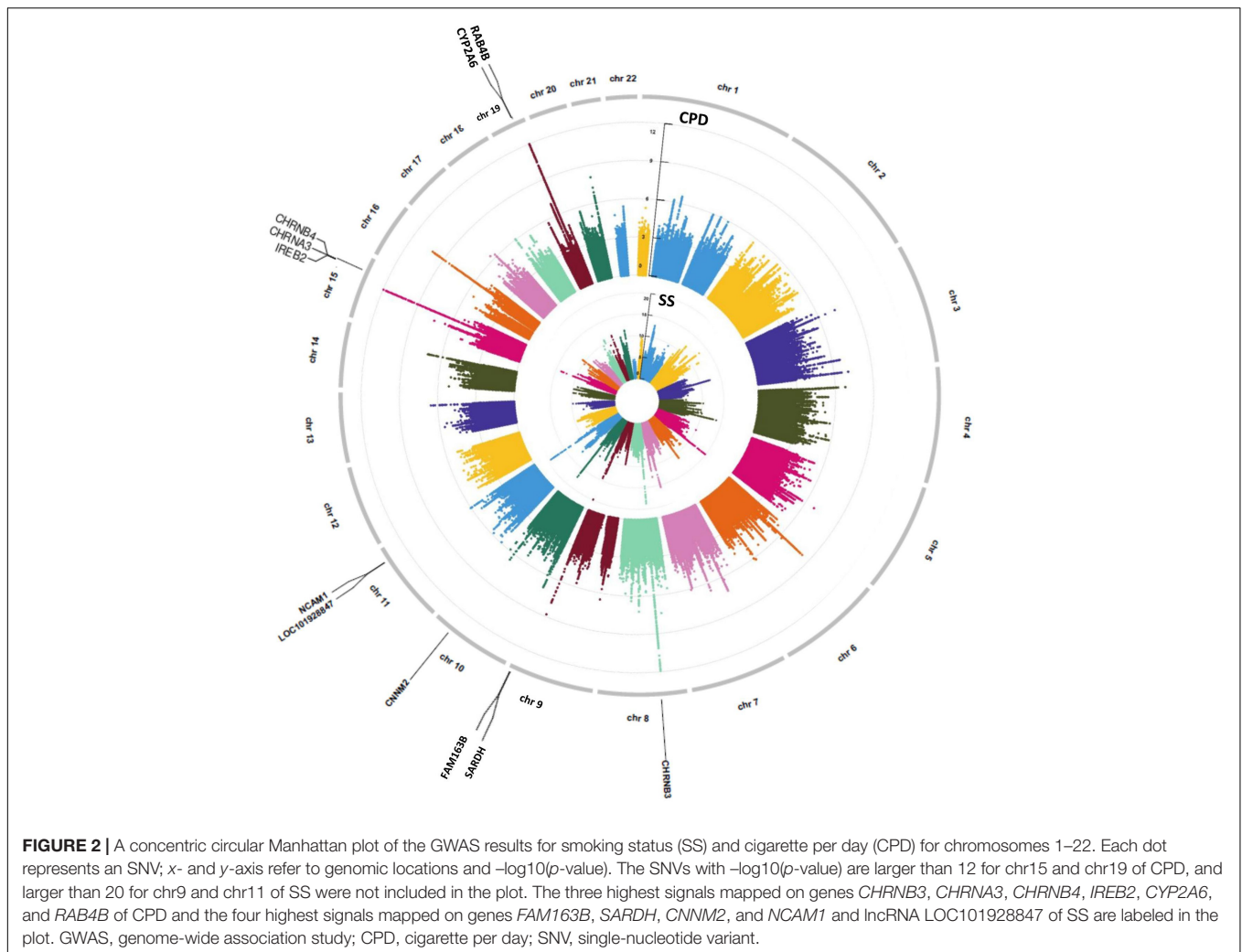
The BIC-based model selection performs exploratory mediation analysis to determine the favored mediation model for each potentially causal variant. We then validated the mediation model selected by carefully checking the causal mediation assumptions (see **Supplementary Methods**). Once the model assumptions are checked, the causal mediation has been established. For model 1, β_{31j} represents the direct effect, $\beta_{1j}\beta_{32}$ represents the indirect/mediation effect, and β_{2j} represents the total effect. For model 2, β_{41j} represents the direct effect, $\beta_{2j}\beta_{42}$ represents the indirect/mediation effect, and β_{1j} represents the total effect. Zhao et al. (2010) classified mediation into three types according to significance and direction of the direct effect when mediation effect is significant: when the direct effect is also significant and has the same sign, the mediation is called a complementary mediation; if they point to the opposite directions, the mediation is called a competitive mediation; lastly, if the direct effect is not significant, the mediation is indirect-only mediation. We followed this classification to interpret the final mediation results for each variant.

All statistical analyses were conducted using R (R Core Team, 2020). An R package "*mediation*" (Tingley et al., 2014) was used for model checking in mediation analysis.

RESULTS

Genome-Wide Association Study and Selection of Smoking Associated Loci

Genome-wide association study were conducted separately for SS and CPD ($N = 293,759$ and $142,202$, respectively). Numerous important smoking behavior associated loci previously reported were reproduced in our study (Gelernter et al., 2006; Thorgeirsson et al., 2008; Keskitalo et al., 2009; Bloom et al., 2014; Bidwell et al., 2015a; Erzurumluoglu et al., 2019; Xu et al., 2020) as highlighted in the circular Manhattan plots (**Figure 2**). Notably, the significant loci identified for each of the two traits have little overlap, implying the different genetic bases of the two nicotine dependence-related smoking phenotypes. The loci associated with SS are mainly located in regions on chromosomes 9, 10, and 11 marked by genes *FAM163B*, *SARDH*, *CNNM2*, and *NCAM1* and non-coding RNA LOC101928847, while loci associated with CPD are located in regions on chromosomes 8, 15, and 19 marked by genes

FIGURE 2 | A concentric circular Manhattan plot of the GWAS results for smoking status (SS) and cigarette per day (CPD) for chromosomes 1–22. Each dot represents an SNV; x- and y-axis refer to genomic locations and –log10(p-value). The SNVs with –log10(p-value) are larger than 12 for chr15 and chr19 of CPD, and larger than 20 for chr9 and chr11 of SS were not included in the plot. The three highest signals mapped on genes *CHRNB3*, *CHRNA3*, *CHRNB4*, *IREB2*, *CYP2A6*, and *RAB4B* of CPD and the four highest signals mapped on genes *FAM163B*, *SARDH*, *CNNM2*, and *NCAM1* and lncRNA LOC101928847 of SS are labeled in the plot. GWAS, genome-wide association study; CPD, cigarette per day; SNV, single-nucleotide variant.

CHRNB3, *CHRNA3*, *IREB2*, and *RAB4B* (**Table 1**). The results validated the gene findings of smoking behaviors in previous GWASs (Ware et al., 2011; Lassi et al., 2016).

Next, we performed causal pathway analysis on the significantly associated loci ($p < 5e\text{-}8$) for SS and CPD. Considering the strong LD among nearby loci, we also included loci in the extended genomic regions by 250 kb both upstream and downstream of the peak regions (**Supplementary Figures 2–7**). This covers genomic regions including a total of 5,828 SNPs for SS and 4,420 SNPs for CPD (**Table 1**). In the causal pathway analysis stage, we focused on participants who have genotype, FA measure, and smoking phenotype data available ($N = 23,624$ for SS and $N = 8,830$ for CPD) and further narrowed down the analysis to pleiotropic variants that were also associated with FA measures (see Methods of Causal pathway analysis Step 1).

Causal Pathway Analysis for Smoking Status

Univariate association analysis found 29 FA measures from various brain regions that show significantly lower FA values

among current smokers than never smokers ($\hat{\beta} < 0$, $p < 0.05$; **Supplementary Table 3**), supporting the findings in previous literature (Savjani et al., 2014; Umene-Nakano et al., 2014; Gray et al., 2020). We proceeded with the causal pathway analysis for each of these 29 FA measures. Only the FA measure in the right tract of the ALIC (ALIC-R) had pleiotropic variants that passed the statistical significance thresholds (overall FDR < 0.15); thus, we continued with this FA measure only in the subsequent steps. We observed 272 SNPs having a pleiotropic effect on SS and ALIC-R (detailed annotation information from FAVOR is included in **Supplementary Table 5**). The significant association between ALIC-R and SS held given the genetic effects of any of the SNPs, implying a vertical pleiotropic relationship (**Supplementary Table 4**), so we no longer proceed with model 0. In comparing the two mediation models of vertical pleiotropy, model 1 where FA mediates the genetic effect on SS is favored. Key mediation assumptions were checked for the chosen model. A majority of those variants (244 out of 272) resided in gene *NCAM1* (**Supplementary Table 5**). The role of *NCAM1* in addiction has been found in recent studies (Bidwell et al., 2015b; Muskiewicz et al., 2018).

TABLE 1 | The selected genomic regions for causal pathway analysis and representative marker genes.

CHR	Regions selected		Numbers of SNPs selected	Numbers of SNPs with GWAS p-value $<$ 1e-8	Representative marker genes	References
	Start bp	End bp				
Cigarette per day (CPD)						
8	42,302,562	42,842,209	1,321	130	CHRNB3	Thorgeirsson et al. (2008); Erzurumluoglu et al. (2019)
15	78,635,394	79,163,637	1,523	1,016	CHRNA3, CHRNB4, IREB2	Keskitalo et al. (2009); Erzurumluoglu et al. (2019)
19	41,033,670	41,552,849	1,576	428	RAB4B, CYP2A6	Thorgeirsson et al. (2008); Bloom et al. (2014), Erzurumluoglu et al. (2019)
Smoking status (SS)						
9	136,192,141	136,724,472	1,889	47	FAM163B, SARDH	Furberg et al. (2010); Xu et al. (2020)
10	104,428,075	105,088,344	1,547	536	CNNM2	Erzurumluoglu et al. (2019); Xu et al. (2020)
11	112,580,002	113,399,158	2,392	252	LOC101928847, NCAM1	Gelernter et al. (2006); Bidwell et al. (2015a), Xu et al. (2020)

SNPs, single-nucleotide polymorphisms; GWAS, genome-wide association study.

Zhu et al. (2016) discussed about two possible explanations for an observed association between genotype and trait, including true causality or strong LD with true causal variants in the same locus. Here, we did not distinguish the 244 variants for true causality vs. linkage and treated all of them as candidate causal variants. **Figure 3** shows how these SNPs impacted SS via ALIC-R. Current smokers carry a significantly larger number of minor allele copies of these SNPs than never smokers on average ($\hat{\beta}_{41} > 0$). The indirect effect of carrying more minor alleles via FA in ALIC-R is in the opposite direction of the direct effect ($\hat{\beta}_2\hat{\beta}_{42} < 0$), thus regarded as competitive mediation (see **Supplementary Table 6** for a complete mediation analysis results).
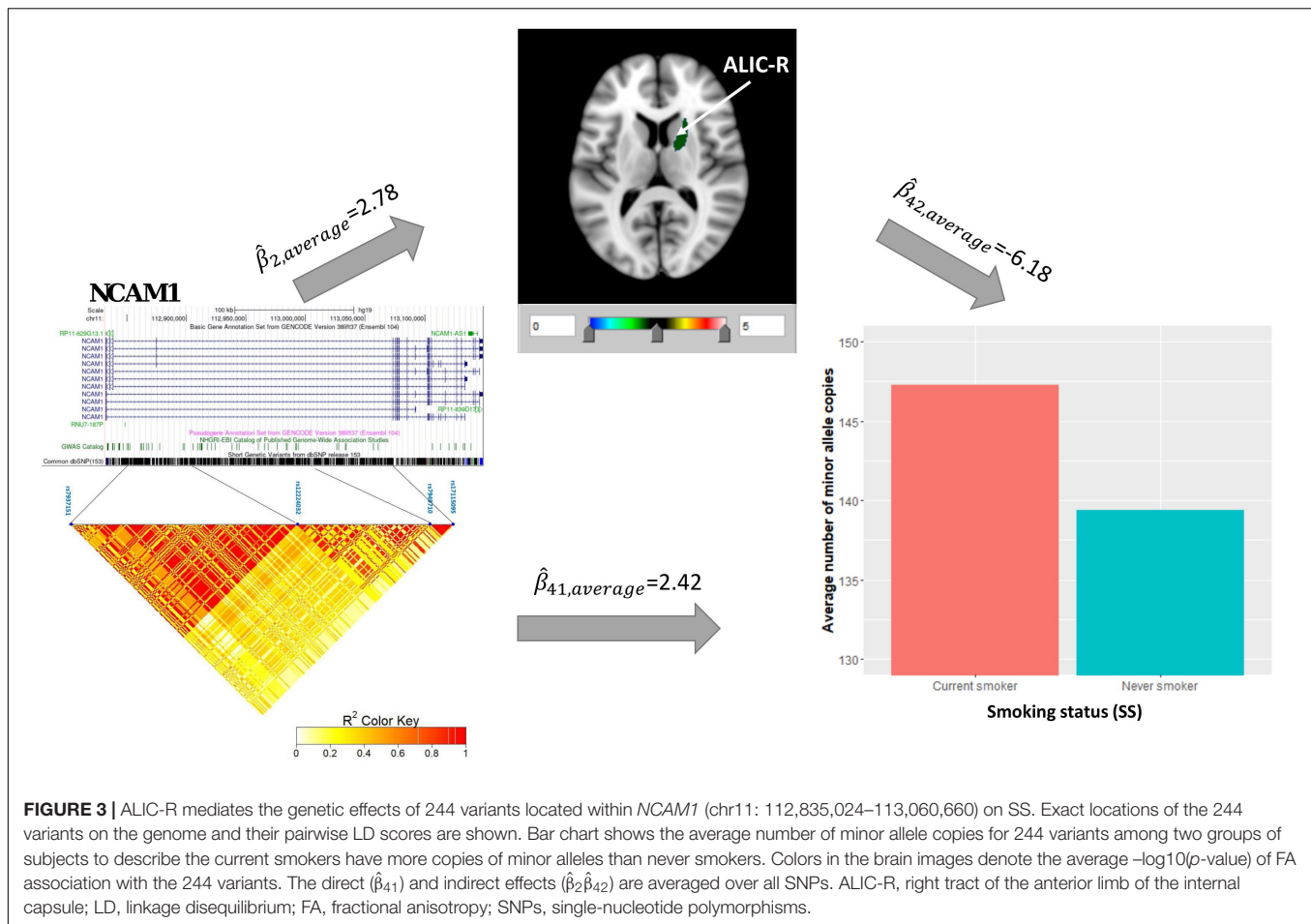
Causal Pathway Discovery for Cigarette per Day

Univariate association found 29 FA measures that show significantly negative association with CPD ($\hat{\beta} < 0$, $p < 0.05$; **Supplementary Table 3**). Among these 29 FA measures, only two FA measures in the ALIC-R and left tract of the PCR (PCR-L) had pleiotropic variants that passed the statistical significance thresholds (overall FDR < 0.15); thus, we continued with these two FA measures only in the subsequent steps. We observed 22 SNPs having a pleiotropic effect on CPD and FA measures ALIC-R and PCR-L (**Supplementary Table 5**). CPD and FA measures are dependent on each other given the genetic effects (**Supplementary Table 4**), implying a vertical pleiotropic relationship. Model 2 (CPD mediates the genetic effect on FA) was chosen as the best mediation model for these 22 SNPs, and the key mediation assumptions were checked for the chosen model. These variants are located in the exonic, intronic, and untranslated regions of gene IREB2 (**Supplementary Table 5**). **Figure 4** shows how these SNPs impacted the two regional FA measures ALIC-R and PCR-L via CPD. The minor alleles

of these SNPs appear to have a protective direct effect on brain structure ($\hat{\beta}_{41} > 0$), while exerting an adverse effect associating with higher CPD ($\hat{\beta}_2\hat{\beta}_{42} < 0$), regarded as competitive mediation (see **Supplementary Table 6** for a complete mediation analysis results).
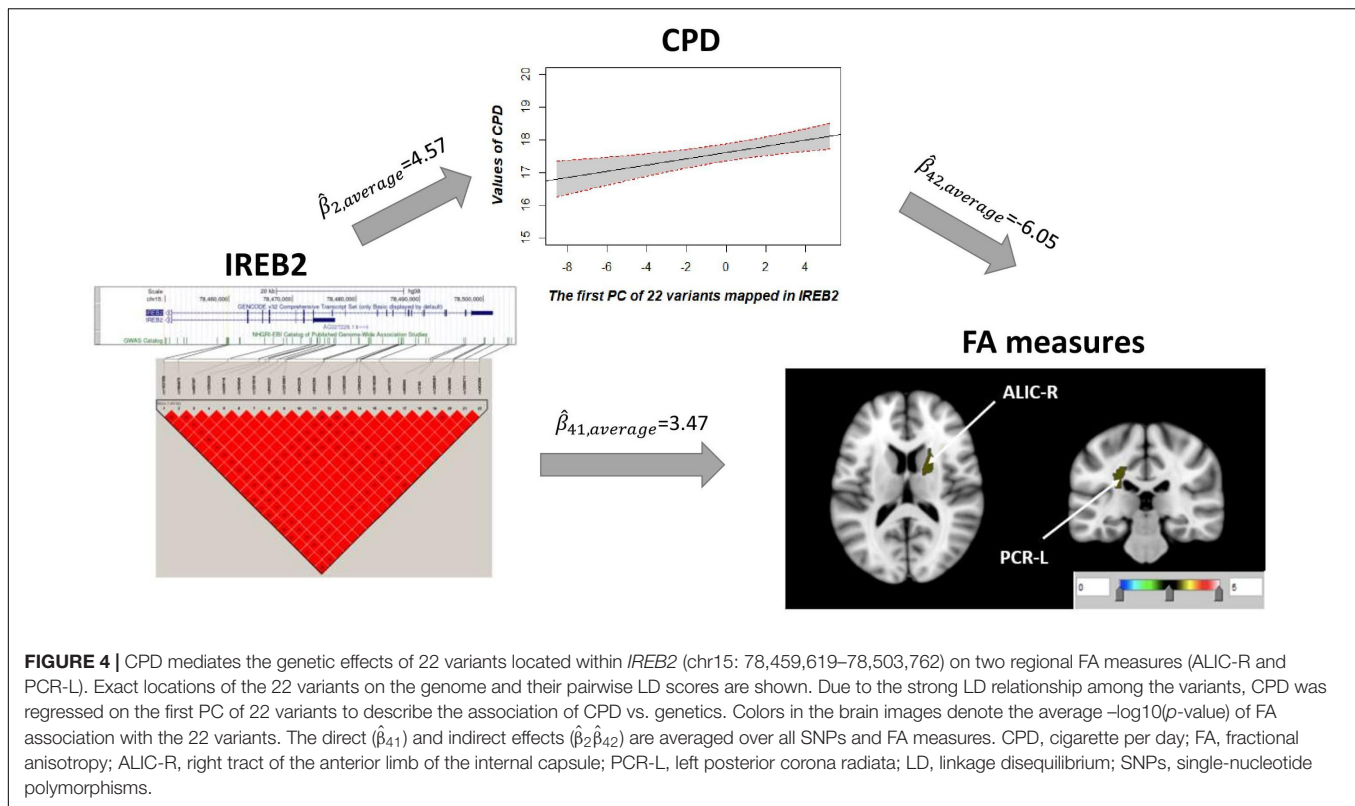
DISCUSSION

In this study, we used novel causative imaging genetics analyses to test neurogenetic mechanisms of nicotine dependence through altered WM integrity. We hypothesized and tested three pleiotropic models to explain the complex causal relationship among genetics, WM integrity, and nicotine dependence. Our GWAS on SS and CPD identified two different sets of associated genetic variants including many reported to be related to smoking in previous large-scale GWASs or meta-analyses (Hardin et al., 2012; Wain et al., 2015; Zeng et al., 2020). We also found smoking (being current smoker or having higher CPD) to be associated with lower WM integrity measured by FA in multiple brain regions. The causal pathway analysis identified 272 pleiotropic SNPs associated with FA in ALIC(R) and SS, and 22 pleiotropic SNPs associated with FA in ALIC(R) and PCR(L) and CPD. These SNPs are mainly located in genes NCAM1 and IREB2. NCAM1 was found to influence risk of nicotine addiction (Gelernter et al., 2006; Muskiewicz et al., 2018). IREB2, which regulates iron mechanism in the cell, was a susceptibility gene for both neurodegeneration and smoking-induced diseases (DeMeo et al., 2009; Cooper et al., 2019). Interestingly, these two sets of SNPs favored different vertical pleiotropic pathways: Gene $\rightarrow$ FA $\rightarrow$ SS vs. Gene $\rightarrow$ CPD $\rightarrow$ FA. The basic genetic components of addiction might have produced a pattern change in WM among smokers, reinforcing the addiction behavior. Chronic severe smoking (reflected in, e.g., CPD) will have negative impact to overall health, which in turn reduces the WM integrity.

FIGURE 3 | ALIC-R mediates the genetic effects of 244 variants located within *NCAM1* (chr11: 112,835,024–113,060,660) on SS. Exact locations of the 244 variants on the genome and their pairwise LD scores are shown. Bar chart shows the average number of minor allele copies for 244 variants among two groups of subjects to describe the current smokers have more copies of minor alleles than never smokers. Colors in the brain images denote the average –log10(*p*-value) of FA association with the 244 variants. The direct ($\hat{\beta}_{41}$) and indirect effects ($\hat{\beta}_2\hat{\beta}_{42}$) are averaged over all SNPs. ALIC-R, right tract of the anterior limb of the internal capsule; LD, linkage disequilibrium; FA, fractional anisotropy; SNPs, single-nucleotide polymorphisms.

We have used an imaging genetics approach to test the neurogenetic mechanisms. Traditional imaging genetics treat neuroimaging as the mediator of the genetic effect on behavior through a unidirectional model (Meyer-Lindenberg and Weinberger, 2006; Bogdan et al., 2017). We use novel causal mediation models to evaluate vertical and horizontal pleiotropy pathways, SNP → behavior → brain vs. SNP → brain → behavior vs. SNP → behavior and SNP → brain, in understanding the neurogenetic mechanism of nicotine addiction behavior. Constraint-based methods and score-based methods are two main categories of conventional causal discovery methods (Glymour et al., 2019). Constraint-based methods identified causal links by conducting conditional independence tests, while score-based methods selected the model with the optimal score from multiple candidate causal models (Spirtes and Zhang, 2016). Here, we proposed to perform conditional independence test to distinguish vertical pleiotropy from horizontal pleiotropy, and we used a BIC score-based method to select the optimal model with the larger score from the two competing vertical pleiotropic models represented by directed acyclic graph (DAG). Such hybrid approach exploited principled ways to combine advantages of both methods and was a computationally efficient strategy to learn the causal structure in a wide range of real-life applications (Wong et al., 2002;

Tsamardinos et al., 2006; Scutari et al., 2018). In our data, such an approach indeed resulted in a DAG with higher likelihood values.

Vertical pleiotropy and horizontal pleiotropy are two competing types of pleiotropy in complex polygenic traits (Paaby and Rockman, 2013). Conventional MR framework treats pleiotropic genetic factors as IV to elucidate how one phenotype (modifiable exposure) causally relates to another phenotype (the outcome) (Davey Smith and Hemani, 2014), but it only considers the vertical pleiotropy with no direct causal link between SNP and outcome (known as exclusion restriction assumption) (Hemani et al., 2018a). We developed a model that evaluates both vertical and horizontal pleiotropy pathways simultaneously and used a rigorous likelihood-based approach to determine the optimal model. We found a significant direct effect of SNPs on the outcomes using our data, which violates the exclusion restriction assumption (Davey Smith and Hemani, 2014). Such extension from the MR framework to tolerate SNPs violating the exclusion restriction assumption of IV analysis has also been seen in recent genetic literature (Bowden et al., 2015; Verbanck et al., 2018). In addition, to establish causality, we also carefully checked the main causal mediation model assumptions of the optimal model.

FIGURE 4 | CPD mediates the genetic effects of 22 variants located within *IREB2* (chr15: 78,459,619–78,503,762) on two regional FA measures (ALIC-R and PCR-L). Exact locations of the 22 variants on the genome and their pairwise LD scores are shown. Due to the strong LD relationship among the variants, CPD was regressed on the first PC of 22 variants to describe the association of CPD vs. genetics. Colors in the brain images denote the average −log10(*p*-value) of FA association with the 22 variants. The direct ($\hat{\beta}_{41}$) and indirect effects ($\hat{\beta}_2 \hat{\beta}_{42}$) are averaged over all SNPs and FA measures. CPD, cigarette per day; FA, fractional anisotropy; ALIC-R, right tract of the anterior limb of the internal capsule; PCR-L, left posterior corona radiata; LD, linkage disequilibrium; SNPs, single-nucleotide polymorphisms.

The biophysical mechanisms that link smoking and lower WM integrity are unknown. Most studies in chronic and heavy smokers reported reduced FA values when compared with those in nonsmokers (Swan and Lessov-Schlaggar, 2007; Hudkins et al., 2010; Kim et al., 2010; Cullen et al., 2011; Gons et al., 2011; Liao et al., 2011; Zhang et al., 2011a,b). Absolute WM FA values are sensitive to many parameters including myelin content, intra-voxel axonal crossing, and axonal fiber density and diameter (Beaulieu, 2002). However, long-term changes in regional FA values were shown to be mainly ($r > 0.8$) driven by changes in regional cerebral myelin concentrations and myelin packing (Song et al., 2003, 2005; Madler et al., 2008). Our findings strongly implicate cerebral WM in the maintenance of this complex addiction and provide genetic targets for further analyses. This finding should encourage future research to examine how changes in WM integrity may or may not contribute to the overall nicotine effects on brain and cognition. Heavy chronic smoking increases the risk of the development of abnormalities in vascular endothelial morphology and function, which may cause cerebral perfusion leading to poorer neurocognition (Pittilo, 2000). Additionally, chronic smoking can impact the vasomotor reactivity/responsivity of the cerebrovascular through upregulation of Ca^{2+} channels and/or modulation of nitric oxide, resulting in the reduction of regional cerebral blood flow (Zubieta et al., 2001; Domino et al., 2004) and the development of WM disease (Liao et al., 1997; Ding et al., 2003; Jeerakathil et al., 2004).

Our study population consists of individuals with white ethnic background, and we pooled the UKBB data from different assessment sites and phases together in the analyses without considering the heterogeneity across sites and phases. Future studies are needed to evaluate the generalization of our conclusion to individuals from other races or ethnicities and to assess whether the conclusions are related to site and phase effects. Our study focused on the most significant variants from GWAS for the causal pathway analysis. Considering the high dimensionality of SNPs and the complex polygenic architecture of both smoking and WM integrity, SNPs with weak signals may not be identified. Further investigation in large independent cohorts is needed to validate current causal analysis results and provide a full picture of the complex genetic architecture of smoking and brain structure, which will in turn improve our understanding of the neurogenetic mechanism of nicotine addiction behavior.

DATA AVAILABILITY STATEMENT

The raw genetic and phenotypic data used in the current study are available from the UK Biobank (UKBB), which can be accessed via https://www.ukbiobank.ac.uk/.

AUTHOR CONTRIBUTIONS

ZY, CM, and SL performed the analysis and wrote the manuscript. TM, SC, and PK supervised the project and took the

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins. 2021.738037/full#supplementary-material

REFERENCES

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L., Griffanti, L., Douaud, G., et al. (2018). Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* 166, 400–424. doi: 10.1016/j.neuroimage.2017.10.034

Baron, R. M., and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173

Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system – a technical review. *NMR Biomed.* 15, 435–455. doi: 10.1002/nbm.782

Benowitz, N. L. (2010). Nicotine addiction. *N. Engl. J. Med.* 362, 2295–2303. doi: 10.1056/NEJMra0809890

Bernardo, J. M., and Smith, A. F. (2009). *Bayesian Theory*. Hoboken, NJ: John Wiley & Sons.

Bidwell, L., McGeary, J., Gray, J., Palmer, R., Knopik, V., and MacKillop, J. (2015a). An initial investigation of associations between dopamine-linked genetic variation and smoking motives in African Americans. *Pharmacol. Biochem. Behav.* 138, 104–110. doi: 10.1016/j.pbb.2015.09.018

Bidwell, L., McGeary, J., Gray, J., Palmer, R., Knopik, V., and MacKillop, J. (2015b). NCAM1-TTC12-ANKK1-DRD2 variants and smoking motives as intermediate phenotypes for nicotine dependence. *Psychopharmacology* 232, 1177–1186. doi: 10.1007/s00213-014-3748-2

Bloom, A. J., Baker, T. B., Chen, L.-S., Breslau, N., Hatsukami, D., Bierut, L. J., et al. (2014). Variants in two adjacent genes, EGLN2 and CYP2A6, influence smoking behavior related to disease risk via different mechanisms. *Hum. Mol. Genet.* 23, 555–561. doi: 10.1093/hmg/ddt432

Bogdan, R., Salmeron, B. J., Carey, C. E., Agrawal, A., Calhoun, V. D., Garavan, H., et al. (2017). Imaging genetics and genomics in psychiatry: a critical review of progress and potential. *Biol. Psychiatry* 82, 165–175.

Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* 44, 512–525. doi: 10.1093/ije/dyv080

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2017). Genome-wide genetic data on~ 500,000 UK Biobank participants. *bioRxiv* [Preprint]. doi: 10.1101/166298

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8

Cooper, M. S., Stark, Z., Lunke, S., Zhao, T., and Amor, D. J. (2019). IREB2-associated neurodegeneration. *Brain* 142:e40. doi: 10.1093/brain/awz183

Cullen, K. R., Wallace, S., Magnotta, V. A., Bockholt, J., Ehrlich, S., Gollub, R. L., et al. (2011). Cigarette smoking and white matter microstructure in schizophrenia. *Psychiatry Res.* 201, 152–158. doi: 10.1016/j.pscychresns.2011.08.010

Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23, R89–R98. doi: 10.1093/hmg/ddu328

DeMeo, D. L., Mariani, T., Bhattacharya, S., Srisuma, S., Lange, C., Litonjua, A., et al. (2009). Integration of genomic and genetic approaches implicates IREB2 as a COPD susceptibility gene. *Am. J. Hum. Genet.* 85, 493–502. doi: 10.1016/j.ajhg.2009.09.004

Denison, D. G., Holmes, C. C., Mallick, B. K., and Smith, A. F. (2002). *Bayesian Methods for Nonlinear Classification and Regression*. New York, NY: John Wiley & Sons.

Ding, J., Nieto, F. J., Beauchamp, N. J., Longstreth, W. Jr., Manolio, T. A., Hetmanski, J. B., et al. (2003). A prospective analysis of risk factors for white matter disease in the brain stem. *Neuroepidemiology* 22, 275–282. doi: 10.1159/000071190

Ding, Y.-S., Fowler, J. S., Logan, J., Wang, G.-J., Telang, F., Garza, V., et al. (2004). 6-[18 F] Fluoro-A-85380, a new PET tracer for the nicotinic acetylcholine receptor: studies in the human brain and in vivo demonstration of specific binding in white matter. *Synapse* 53, 184–189. doi: 10.1002/syn.20051

Domino, E. F., Ni, L., Xu, Y., Koeppe, R. A., Guthrie, S., and Zubieta, J.-K. (2004). Regional cerebral blood flow and plasma nicotine after smoking tobacco cigarettes. *Prog. Neuro Psychopharmacol. Biol. Psychiatry* 28, 319–327. doi: 10.1016/j.pnpbp.2003.10.011

Donny, E. C., Griffin, K. M., Shiffman, S., and Sayette, M. A. (2008). The relationship between cigarette use, nicotine dependence, and craving in laboratory volunteers. *Nicotine Tob. Res.* 10, 447–455. doi: 10.1080/14622200801901906

Erzurumluoglu, A. M., Liu, M., Jackson, V. E., Barnes, D. R., Datta, G., Melbourne, C. A., et al. (2019). Meta-analysis of up to 622,409 individuals identifies 40 novel smoking behaviour associated genetic loci. *Mol. Psychiatry* 25, 2392–2409. doi: 10.1038/s41380-018-0313-0

Feng, Q., Abraham, J., Feng, T., Song, Y., Elston, R. C., and Zhu, X. (2009). A method to correct for population structure using a segregation model. *BMC Proc.* 3(Suppl. 7):S104. doi: 10.1186/1753-6561-3-S7-S104

Fujita, M., Ichise, M., Van Dyck, C. H., Zoghbi, S. S., Tamagnan, G., Mukhin, A. G., et al. (2003). Quantification of nicotinic acetylcholine receptors in human brain using [123 I] 5-IA-85380 SPET. *Eur. J. Nucl. Med. Mol. Imaging* 30, 1620–1629. doi: 10.1007/s00259-003-1320-0

Furberg, H., Kim, Y., Dackor, J., Boerwinkle, E., Franceschini, N., Ardissino, D., et al. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* 42, 441–447. doi: 10.1038/ng.571

Gelernter, J., Yu, Y., Weiss, R., Brady, K., Panhuysen, C., Yang, B.-Z., et al. (2006). Haplotype spanning TTC12 and ANKK1, flanked by the DRD2 and NCAM1 loci, is strongly associated to nicotine dependence in two distinct American populations. *Hum. Mol. Genet.* 15, 3498–3507. doi: 10.1093/hmg/ddl426

Glymour, C., Zhang, K., and Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Front. Genet.* 10:524. doi: 10.3389/fgene.2019.00524

Gogliettino, A. R., Potenza, M. N., and Yip, S. W. (2016). White matter development and tobacco smoking in young adults: a systematic review with recommendations for future research. *Drug Alcohol Depend.* 162, 26–33. doi: 10.1016/j.drugalcdep.2016.02.015

Gons, R. A., van Norden, A. G., de Laat, K. F., van Oudheusden, L. J., van Uden, I. W., Zwiers, M. P., et al. (2011). Cigarette smoking is associated with reduced microstructural integrity of cerebral white matter. *Brain* 134, 2116–2124. doi: 10.1093/brain/awr145

Gray, J. C., Thompson, M., Bachman, C., Owens, M. M., Murphy, M., and Palmer, R. (2020). Associations of cigarette smoking with gray and white matter in the

UK Biobank. *Neuropsychopharmacology* 45, 1215–1222. doi: 10.1038/s41386-020-0630-2

Hardin, M., Zielinski, J., Wan, E. S., Hersh, C. P., Castaldi, P. J., Schwinder, E., et al. (2012). CHRNA3/5, IREB2, and ADCY2 are associated with severe chronic obstructive pulmonary disease in Poland. *Am. J. Respir. Cell Mol. Biol.* 47, 203–208. doi: 10.1165/rcmb.2012-0011OC

Heckerman, D., Meek, C., and Cooper, G. (1999). "A Bayesian approach to causal discovery," in *Computation, Causation, and Discovery*, Vol. 19, eds C. Glymour, and G. F. Cooper (Menlo Park, CA: MIT Press), 141–166.

Hemani, G., Bowden, J., and Davey Smith, G. (2018a). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* 27, R195–R208. doi: 10.1093/hmg/ddy163

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., et al. (2018b). The MR-Base platform supports systematic causal inference across the human phenome. *elife* 7:e34408. doi: 10.7554/eLife.34408.012

Hillmer, A., Wooten, D., Moirano, J., Slesarev, M., Barnhart, T., Engle, J., et al. (2011). Specific α4β2 nicotinic acetylcholine receptor binding of [F-18] nifene in the rhesus monkey. *Synapse* 65, 1309–1318. doi: 10.1002/syn.20965

Hudkins, M., O'Neill, J., Tobias, M. C., Bartzokis, G., and London, E. D. (2010). Cigarette smoking and white matter microstructure. *Psychopharmacology* 221, 285–295. doi: 10.1007/s00213-011-2621-9

James, L. R., and Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *J. Appl. Psychol.* 69, 307–321. doi: 10.1037/0021-9010.69.2.307

Jeerakathil, T., Wolf, P. A., Beiser, A., Hald, J. K., Au, R., Kase, C. S., et al. (2004). Cerebral microbleeds: prevalence and associations with cardiovascular risk factors in the Framingham study. *Stroke* 35, 1831–1835. doi: 10.1161/01.STR.0000131809.35202.1b

Jordan, D. M., Verbanck, M., and Do, R. (2019). HOPS: a quantitative score reveals pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. *Genome Biol.* 20:222. doi: 10.1186/s13059-019-1844-7

Judd, C. M., and Kenny, D. A. (1981). Process analysis estimating mediation in treatment evaluations. *Eval. Rev.* 5, 602–619.

Kang, S. J., Larkin, E. K., Song, Y., Barnholtz-Sloan, J., Baechle, D., Feng, T., et al. (2009). Assessing the impact of global versus local ancestry in association studies. *BMC Proc.* 3(Suppl. 7):S107.

Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795. doi: 10.1080/01621459.1995.10476572

Kendler, K. S., Neale, M., Sullivan, P., Corey, L., Gardner, C., and Prescott, C. (1999). A population-based twin study in women of smoking initiation and nicotine dependence. *Psychol. Med.* 29, 299–308.

Keskitalo, K., Broms, U., Heliövaara, M., Ripatti, S., Surakka, I., Perola, M., et al. (2009). Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNB4) on chromosome 15. *Hum. Mol. Genet.* 18, 4007–4012. doi: 10.1093/hmg/ddp322

Kim, S. H., Yun, C. H., Lee, S. Y., Choi, K. H., Kim, M. B., and Park, H. K. (2010). Age-dependent association between cigarette smoking on white matter hyperintensities. *Neurol. Sci.* 33, 45–51. doi: 10.1007/s10072-011-0617-1

Kochunov, P., Du, M., Moran, L., Sampath, H., Wijtenburg, A., Yang, Y., et al. (2013). Acute nicotine administration effects on fractional anisotropy of cerebral white matter and associated attention performance. *Front. Pharmacol.* 4:117. doi: 10.3389/fphar.2013.00117

Kochunov, P., Thompson, P. M., Lancaster, J. L., Bartzokis, G., Smith, S., Coyle, T., et al. (2007). Relationship between white matter fractional anisotropy and other indices of cerebral health in normal aging: tract-based spatial statistics study of aging. *Neuroimage* 35, 478–487. doi: 10.1016/j.neuroimage.2006.12.021

Lang, P. M., Burgstahler, R., Sippel, W., Irnich, D., Schlotter-Weigel, B., and Grafe, P. (2003). Characterization of neuronal nicotinic acetylcholine receptors in the membrane of unmyelinated human C-fiber axons by in vitro studies. *J. Neurophysiol.* 90, 3295–3303. doi: 10.1152/jn.00512.2003

Lassi, G., Taylor, A. E., Timpson, N. J., Kenny, P. J., Mather, R. J., Eisen, T., et al. (2016). The CHRNA5–A3–B4 gene cluster and smoking: from discovery to therapeutics. *Trends Neurosci.* 39, 851–861. doi: 10.1016/j.tins.2016.10.005

Li, X., Li, Z., Zhou, H., Gaynor, S. M., Liu, Y., Chen, H., et al. (2020). Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat. Genet.* 52, 969–983. doi: 10.1038/s41588-020-0676-4

Liao, D., Cooper, L., Cai, J., Toole, J., Bryan, N., Burke, G., et al. (1997). The prevalence and severity of white matter lesions, their relationship with age, ethnicity, gender, and cardiovascular disease risk factors: the ARIC Study. *Neuroepidemiology* 16, 149–162. doi: 10.1159/000368814

Liao, Y., Tang, J., Deng, Q., Deng, Y., Luo, T., Wang, X., et al. (2011). Bilateral fronto-parietal integrity in young chronic cigarette smokers: a diffusion tensor imaging study. *PLoS One* 6:e26460. doi: 10.1371/journal.pone.0026460

MacKinnon, D. P., Fairchild, A. J., and Fritz, M. S. (2007). Mediation analysis. *Annu. Rev. Psychol.* 58, 593–614. doi: 10.1146/annurev.psych.58.110405.085542

Madler, B., Drabycz, S. A., Kolind, S. H., Whittall, K. P., and MacKay, A. L. (2008). Is diffusion anisotropy an accurate monitor of myelination? Correlation of multicomponent T2 relaxation and diffusion tensor anisotropy in human brain. *Magn. Reson. Imaging* 26, 874–888. doi: 10.1016/j.mri.2008.01.047

McBride, P. E. (1992). The health consequences of smoking: cardiovascular diseases. *Med. Clin. North Am.* 76, 333–353. doi: 10.1016/S0025-7125(16)30356-X

Meyer-Lindenberg, A., and Weinberger, D. R. (2006). Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nat. Rev. Neurosci.* 7, 818–827. doi: 10.1038/nrn1993

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., et al. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* 19, 1523–1536. doi: 10.1038/nn.4393

Muskiewicz, D. E., Uhl, G. R., and Hall, F. S. (2018). The role of cell adhesion molecule genes regulating neuroplasticity in addiction. *Neural Plast.* 2018:9803764. doi: 10.1155/2018/9803764

Paaby, A. B., and Rockman, M. V. (2013). The many faces of pleiotropy. *Trends Genet.* 29, 66–73.

Perkins, K. A. (1999). Nicotine self-administration. *Nicotine Tob. Res.* 1(Suppl. 2), S133–S137.

Pimlott, S., Piggott, M., Owens, J., Greally, E., Court, J., Jaros, E., et al. (2004). Nicotinic acetylcholine receptor distribution in Alzheimer's disease, dementia with Lewy bodies, Parkinson's disease, and vascular dementia: in vitro binding study using 5-[125 I]-A-85380. *Neuropsychopharmacology* 29, 108–116. doi: 10.1038/sj.npp.1300302

Pittilo, M. (2000). Cigarette smoking, endothelial injury and cardiovascular disease. *Int. J. Exp. Pathol.* 81, 219–230. doi: 10.1046/j.1365-2613.2000.00162.x

Price, A. L., Zaitlen, N. A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* 11, 459–463. doi: 10.1038/nrg2813

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/.

Savjani, R. R., Velasquez, K. M., Thompson-Lake, D. G. Y., Baldwin, P. R., Eagleman, D. M., De La Garza, I. I. R., et al. (2014). Characterizing white matter changes in cigarette smokers via diffusion tensor imaging. *Drug Alcohol Depend.* 145, 134–142. doi: 10.1016/j.drugalcdep.2014.10.006

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* 6, 461–464. doi: 10.1214/aos/1176344136

Scutari, M., Graafland, C. E., and Gutiérrez, J. M. (2018). "Who learns better bayesian network structures: constraint-based, score-based or hybrid algorithms?" in *Proceedings of the International Conference on Probabilistic Graphical Models, PMLR*. Available online at: http://proceedings.mlr.press/v72/scutari18a.html

Song, S. K., Sun, S. W., Ju, W. K., Lin, S. J., Cross, A. H., and Neufeld, A. H. (2003). Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. *Neuroimage* 20, 1714–1722. doi: 10.1016/j.neuroimage.2003.07.005

Song, S. K., Yoshino, J., Le, T. Q., Lin, S. J., Sun, S. W., Cross, A. H., et al. (2005). Demyelination increases radial diffusivity in corpus callosum of mouse brain. *Neuroimage* 26, 132–140. doi: 10.1016/j.neuroimage.2005.01.028

Spirtes, P., and Zhang, K. (2016). Causal discovery and inference: concepts and recent methodological advances. *Appl. Inform.* 3:3. doi: 10.1186/s40535-016-0018-x

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12:e1001779. doi: 10.1371/journal.pmed.1001779

Swan, G. E., and Lessov-Schlaggar, C. N. (2007). The effects of tobacco smoke and nicotine on cognition and the brain. *Neuropsychol. Rev.* 17, 259–273. doi: 10.1007/s11065-007-9035-9

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., et al. (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* 8, 153–182. doi: 10.1007/s11682-013-9269-5

Thorgeirsson, T. E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K. P., et al. (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 452, 638–642. doi: 10.1038/nature06846

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). Mediation: R package for causal mediation analysis. *J. Stat. Softw.* 59, 1–38. doi: 10.18637/jss.v059.i05

True, W. R., Xian, H., Scherrer, J. F., Madden, P. A., Bucholz, K. K., Heath, A. C., et al. (1999). Common genetic vulnerability for nicotine and alcohol dependence in men. *Arch. Gen. Psychiatry* 56, 655–661. doi: 10.1001/archpsyc.56.7.655

Tsamardinos, I., Brown, L. E., and Aliferis, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Mach. Learn.* 65, 31–78.

Tyler, A. L., Asselbergs, F. W., Williams, S. M., and Moore, J. H. (2009). Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. *Bioessays* 31, 220–227.

UK Biobank (2007). UK Biobank: Protocol for a Large-Scale Prospective Epidemiological Resource. Available Online at: https://www.nature.com/articles/s41586-018-0579-z

Umene-Nakano, W., Yoshimura, R., Kakeda, S., Watanabe, K., Hayashi, K., Nishimura, J., et al. (2014). Abnormal white matter integrity in the corpus callosum among smokers: tract-based spatial statistics. *PLoS One* 9:e87890. doi: 10.1371/journal.pone.0087890

Verbanck, M., Chen, C.-Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50, 693–698. doi: 10.1038/s41588-018-0099-7

Vink, J. M., Willemsen, G., and Boomsma, D. I. (2005). Heritability of smoking initiation and nicotine dependence. *Behav. Genet.* 35, 397–406. doi: 10.1007/s10519-004-1327-8

Wain, L. V., Shrine, N., Miller, S., Jackson, V. E., Ntalla, I., Artigas, M. S., et al. (2015). Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir. Med.* 3, 769–781. doi: 10.1016/S2213-2600(15)00283-0

Ware, J. J., van den Bree, M. B., and Munafò, M. R. (2011). Association of the CHRNA5-A3-B4 gene cluster with heaviness of smoking: a meta-analysis. *Nicotine Tob. Res.* 13, 1167–1175. doi: 10.1093/ntr/ntr118

Warren, H. R., Evangelou, E., Cabrera, C. P., Gao, H., Ren, M., Mifsud, B., et al. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* 49, 403–415. doi: 10.1038/ng.3768

Wong, M. L., Lee, S. Y., and Leung, K. S. (2002). "A hybrid approach to discover Bayesian networks from databases using evolutionary programming," in *Proceedings of the 2002 IEEE International Conference on Data Mining,* (Piscataway, NJ: IEEE).

Xu, K., Li, B., McGinnis, K. A., Vickers-Smith, R., Dao, C., Sun, N., et al. (2020). Genome-wide association study of smoking trajectory and meta-analysis of smoking status in 842,000 individuals. *Nat. Commun.* 11:5302. doi: 10.1038/s41467-020-18489-3

Zeng, Q., Chen, Q., Zou, D., Guo, R., Xiao, D., Jiang, S., et al. (2020). Different associations between the IREB2 variants and chronic obstructive pulmonary disease susceptibility. *Front. Genet.* 11:598053. doi: 10.3389/fgene.2020.598053

Zhang, X., Salmeron, B. J., Ross, T. J., Geng, X., Yang, Y., and Stein, E. A. (2011a). Factors underlying prefrontal and insula structural alterations in smokers. *Neuroimage* 54, 42–48. doi: 10.1016/j.neuroimage.2010.08.008

Zhang, X., Stein, E. A., and Hong, L. E. (2010). Smoking and schizophrenia independently and additively reduce white matter integrity between striatum and frontal cortex. *Biol. Psychiatry* 68, 674–677. doi: 10.1016/j.biopsych.2010.06.018

Zhang, X., Stein, E. A., and Hong, L. E. (2011b). Smoking and schizophrenia independently and additively reduce white matter integrity between striatum and frontal cortex. *Biol. Psychiatry* 68, 674–677.

Zhao, X., Lynch, J. G. Jr., and Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *J. Consum. Res.* 37, 197–206. doi: 10.1086/651257

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481–487. doi: 10.1038/ng.3538

Ziedonis, D., Hitsman, B., Beckham, J. C., Zvolensky, M., Adler, L. E., Audrain-McGovern, J., et al. (2008). Tobacco use and cessation in psychiatric disorders: National Institute of Mental Health report. *Nicotine Tob. Res.* 10, 1691–1715. doi: 10.1080/14622200802443569

Zubieta, J.-K., Lombardi, U., Minoshima, S., Guthrie, S., Ni, L., Ohl, L. E., et al. (2001). Regional cerebral blood flow effects of nicotine in overnight abstinent smokers. *Biol. Psychiatry* 49, 906–913. doi: 10.1016/S0006-3223(00)01070-2