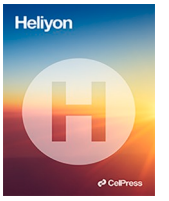
Heliyon 10 (2024) e35649

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

A comprehensive analysis of the health effects associated with smoking in the largest population using UK Biobank genotypic and phenotypic data

Zixun Lin [a,d,1], Jiayi Xiong [a,1], Jiaqi Yang [a], Yuanfeng Huang [a], Jinchen Li [a,b,c,e], Guihu Zhao [a], Bin Li [a,*]

[a] The Joint Institute of Smoking and Health & Bioinformatics Centre, National Clinical Research Centre for Geriatric Disorders, Department of Geriatrics, Xiangya Hospital, Central South University, Changsha, Hunan, 410008, China
[b] Centre for Medical Genetics & Hunan Key Laboratory, School of Life Sciences, Central South University, Changsha, Hunan, 410008, China
[c] Department of Neurology, Xiangya Hospital, Central South University, Changsha, Hunan, 410008, China
[d] Xiangya School of Medicine, Central South University, Changsha, Hunan, 410013, China
[e] Bioinformatics Centre, Furong Laboratory, Changsha, Hunan, 410008, China

ARTICLE INFO

Keywords:
Smoking
Mendelian randomization
Phenome-wide association study
Disease causality

ABSTRACT

Background: Smoking is a widespread behavior, while the relationship between smoking and various diseases remains a topic of debate.
Objective: We conducted analysis to further examine the identified associations and assess potential causal relationships.
Methods: We utilized seven single nucleotide polymorphisms (SNPs) known to be linked to smoking extracting genotype data from the UK Biobank, a large-scale biomedical repository encompassing comprehensive health-related and genetic information of European descent. Phenome-wide association study (PheWAS) analysis was conducted to map the association of genetically predicted smoking status with 1,549 phenotypes. The associations identified in the PheWAS were then meticulously examined through two-sample Mendelian randomization (MR) analysis, utilizing data from the UK Biobank (n = 487,365) and the Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) (n = 337,334). This approach allowed us to comprehensively characterize the links between smoking and disease patterns.
Results: The PheWAS analysis produced 34 phenotypes that demonstrated significant associations with smoking (P = 0.05/1460). Importantly, sickle cell anemia and type 2 diabetes exhibited the most significant SNPs (both 85.71% significant SNPs). Furthermore, the MR analyses provided compelling evidence supporting causal associations between smoking and the risk of following diseases: obstructive chronic bronchitis (IVW: Beta = 0.48, 95% confidence interval (CI) 0.36-0.61, P = 1.62×10^{-13}), cancer of the bronchus (IVW: Beta = 0.92, 95% CI 0.68-1.17, P = $2.02\times10^{-13)}$, peripheral vascular disease (IVW: Beta = 1.09, 95% CI 0.71-1.46, P = 1.63×10^{-8}), emphysema (IVW: Beta = 1.63, 95% CI 0.90-2.36, P = 1.29×10^{-5}), pneumococcal pneumonia (IVW: Beta = 0.30, 95% CI 0.11-0.49, P = 1.60×10^{-3}), chronic airway obstruction (IVW: Beta =

* Corresponding author.
E-mail addresses: 8301210409@csu.edu.cn (Z. Lin), xiongjy@csu.edu.cn (J. Xiong), 228111036@csu.edu.cn (J. Yang), 2025512649@qq.com (Y. Huang), lijinchen@csu.edu.cn (J. Li), ghzhao@csu.edu.cn (G. Zhao), lebin001@csu.edu.cn (B. Li).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e35649
Received 10 March 2024; Received in revised form 29 July 2024; Accepted 1 August 2024
Available online 2 August 2024

0.83, 95% CI 0.30-1.36, P = 2.00×10^{-3}) and type 2 diabetes (IVW: Beta = 0.53, 95% CI 0.16-0.90, P = 5.08×10^{-3}).
Conclusion: This study affirms causal relationships between smoking and obstructive chronic bronchitis, cancer of the bronchus, peripheral vascular disease, emphysema, pneumococcal pneumonia, chronic airway obstruction, type 2 diabetes, in the European population. These findings highlight the broad health impacts of smoking and support smoking cessation efforts.

1. Background

Smoking, a widespread habit, is a significant threat that contributes to many chronic diseases and deaths [1]. According to the World Health Organization, the global populace of smokers surpasses one billion, with an alarming estimate of eight million smoking-related deaths annually [2]. While extensive observational data has persistently associated smoking with respiratory, cardiovascular, and neurological disorders, other studies have questioned these connections [3–5]. It is worth noting that one study revealed that smoking could reduce the occurrence of certain respiratory conditions like allergic asthma [6]. Conversely, an alternative study brought to light that smokers exhibit an adjusted odds ratio (OR) of 1.33 for asthma compared to their non-smoking counterparts, signifying an augmented risk [7]. Furthermore, recent findings have shown that cigarette smoke exacerbates asthma through the induction of lymphocytes [8]. Intriguingly, smoking has also emerged as a protective factor against endometrial cancer and Parkinson's disease, suggesting that the ramifications of smoking on clinical ailments are both comprehensive and intricate [9,10]. To fully grasp the complex impacts of smoking on these illnesses and elucidate possible underlying causes, a thorough examination of condition-related phenotypic clusters is crucial. These groups effectively categorize diseases based on their distinct clinical characteristics, symptoms, and underlying biological pathways, thereby facilitating a nuanced exploration of smoking's impact on diverse disease processes and enabling the identification of vulnerable subpopulations as well as potential shared mechanisms between smoking and diseases.

Although previous studies have partially revealed the association between smoking and numerous diseases from different perspectives, however, these studies often face limitations such as confounding variables, reverse causality, and the inability to establish definitive causal relationships. To overcome the limitations faced by previous researches, our study employs both Phenome-wide association studies (PheWAS) and Mendelian randomization (MR) methodologies. Using PheWAS is an effective method to explore the complex relationships between genetic variants and multiple phenotypes, ultimately leading to the ascertainment of potential relationships [11]. MR, a widely embraced methodology, serves as an instrumental tool to scrutinize the identified associations and assess potential causal relationships [12]. MR allows us to leverage genetic variation to estimate causal effects, thus avoiding confounding influences and reverse causality. By harnessing the natural experiment provided by genetic variation, MR proffers a robust avenue to estimate causal effects while artfully sidestepping the confounding influences and the perils of reverse causality. Consequently, MR emerges as a reliable and steadfast tool, fostering the facilitation of credible causal inference in the realm of research endeavors.

To holistically explore the profound impact of smoking on the array of disorders, we harnessed a substantial corpus of data derived from the UK Biobank. Subsequently, by deploying PheWAS and MR, we fervently anticipate gleaning invaluable insights that can potentially engender transformative advances in the arenas of disease prevention and treatment strategies. We hope this study will provide helpful reference for public health policies and clinical practices to better address smoking-related health issues.

2. Methods

2.1. Study population and research design

To conduct our analysis, we accessed the UK Biobank, an expansive and prospective repository of genetic and health information from participants aged between 40 and 69 years from across the UK in 2006–2010, and employed data from 487,365 participants of European decent [13]. The characteristics of participants are shown in Table 1. Rigorous quality control measures were diligently implemented in consonance with the established protocols [14]. In this study, PheWAS was used to identify the binary phenotypic

Table 1
Characteristics of participants in the UK Biobank (N = 487,365).

Listings	Level	Overall
Number		487,365
Age (mean (SD))		56.54 (8.09)
BMI (mean (SD))		26.82 (5.07)
Sex (%)	Female	264138 (54.2)
	Male	223026 (45.8)
Current tobacco smoking (%)	Prefer not to answer	405 (0.1)
	No	435056 (89.4)
	Yes, on most or all days	37898 (7.8)
	Only occasionally	13286 (2.7)

traits that bear intricate associations with smoking in the full phenotype scale. Subsequently, MR analysis were conducted to reveal potential causal relationships between smoking and the identified phenotypes. All analyses were conducted in R (Version 4.2.2). Fig. 1 shows the work flow of the present research.

2.2. Phenome-wide association study

PheWAS is a valuable methodology that combines electronic health records (EHRs) and DNA repositories to identify phenotypic correlations with specific SNPs [11]. PheWAS has demonstrated its robust capability in systematically exploring a wide range of phenotypes and identifying novel associations with genetic variants. Binary traits in PheWAS are coded by PheCODE, an ingenious phenotypic coding system derived from the International Classification of Diseases (ICD-10) codes [15]. We searched for articles on the PubMed website according to the following search strategy: ("Smoking" [Mesh] OR "Tobacco Use Disorder" [Mesh]) AND "Polymorphism, Single Nucleotide" [Mesh], and seven frequent-appeared SNPs were detected (Table 2). These SNPs include rs16969968 [16], rs13329271 [17], rs1051730 [18], rs12595538 [18], rs8032771 [18], rs6495308 [18], and rs7937 [19].

To investigate their relationship with smoking, we extracted the genotype data for these seven SNPs, as well as binary clinical data from 1,549 phenotypes. Binary traits with fewer than 20 cases or controls were excluded. The genotype data for this study were meticulously gleaned through the utilization of the UK BiLEVE Axiom array and the UK Biobank Axiom array of Affymetrix chips, with reference genome GRCh37 [20]. In this study, PheWAS analysis was conducted based on R package "PheWAS" (Version 0.99.5.5). Bonferroni correction method was used to calculate adjusted P value. This approach ensures that this analysis focuses on robust associations while accounting for sample size limitations in specific traits.

2.3. Mendelian randomization study

2.3.1. Included traits

For MR analysis, the exposure variable was extracted from the genome-wide association study (GWAS) dataset, which contained smoking data from 337,334 participants of European descent on Cigarettes per Day from the GWAS and Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) [21]. The GWAS ID of the dataset on IEU OpenGWAS (https://gwas.mrcieu.ac.uk/) is ieu-b-25. Binary traits that exhibited positive associations in PheWAS at least three times were selected as outcome variables for MR analysis. To identify relevant GWAS for these outcome variables, we chose a biobank-scale study that employed an improved analytical method called the generalized linear mixed model, which could enhance data calibration and analysis efficiency, providing a systematic and comprehensive high-quality GWAS dataset that was downloadable [22]. The data source for the model were derived from 2,989 binary phenotypes in the UK Biobank with 456,348 individuals of European ancestry and 11,842,647 imputed variants.

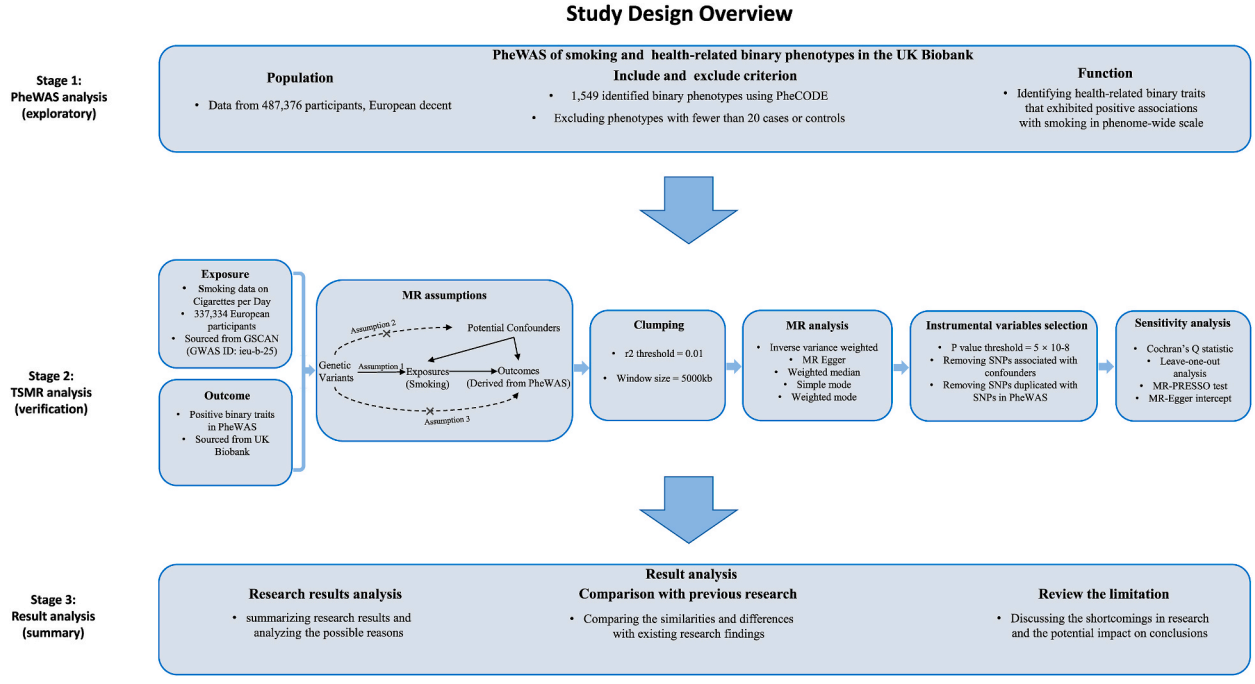


Fig. 1. Research design overview. PheWAS, Phenome-wide association study; MR, Mendelian randomization; GSCAN, GWAS and Sequencing Consortium of Alcohol and Nicotine use.

Table 2
The sources and basic information of SNPs for PheWAS.

Rsid	Chr:pos (GRCh38)	Gene	PMID
rs16969968	15:78590583	CHRNA5	18519524
rs13329271	15:78621888	CHRNA3	31089300
rs1051730	15:78601997	CHRNA3	18385738
rs12595538	15:78862111	MORF4L1	23536134
rs8032771	15:78833717	MORF4L1	
rs6495308	15:78615314	CHRNA3	
rs7937	19:40796801	RAB4B	

2.3.2. MR analysis

To ensure the validity of our Mendelian randomization (MR) analysis, we adhered to three core assumptions: (1) the genetic variants are associated with the exposure (smoking), (2) the variants are not associated with any confounders, and (3) the variants affect the outcome only through the exposure. If an instrumental variable did not appear in the outcome GWAS, it was excluded from the analysis to maintain the integrity of these assumptions.

For Assumption 1, SNPs with $P < 5 \times 10^{-8}$ were served as instrumental variables for MR analysis. In addition, We excluded SNPs located within a 5000-kb range upstream and downstream of the most significant SNP to minimize linkage disequilibrium and avoid including SNPs that might be correlated ($r2 > 0.01$ in the 1000 Genomes European data). For Assumption 2, PhenoScanner (http://

Table 3
PheWAS results showcase, including 34 diseases and the sum of positive results in reverse order.

Classification	Diseases	rs16969968	rs13329271	rs12595538	rs8032771	rs6495308	rs1051730	rs7937
Hematological system diseases	sickle-cell anemia	+	+	+	−	+	+	+
	other hemoglobinopathies	+	+	−	−	+	−	+
	secondary malignancy of lymph nodes	+	−	−	−	−	+	−
Psychiatric diseases	tobacco use disorder	+	+	−	−	+	+	+
	other mental disorder	−	+	−	−	+	−	+
Dermatologic diseases	other non-epithelial cancer of skin	−	−	−	−	+	−	−
	cellulitis and abscess of foot, toe	−	−	−	−	+	−	−
	cellulitis and abscess of arm/hand	−	−	−	−	+	−	−
	cellulitis and abscess of leg, except foot	−	−	−	−	+	−	−
Respiratory diseases	chronic airway obstruction	+	+	−	−	+	+	+
	obstructive chronic bronchitis	+	+	−	−	+	+	+
	cancer of bronchus; lung	+	+	−	+	+	+	−
	emphysema	+	+	−	−	+	+	−
	pneumococcal pneumonia	+	+	−	−	+	−	−
	respiratory insufficiency	+	−	−	−	−	+	−
	cancer within the respiratory system	+	−	−	−	−	+	−
	respiratory failure	+	−	−	−	−	+	−
	pleural effusion; pleurisy	−	−	−	−	+	−	−
	abnormal findings examination of lungs	−	−	−	−	−	+	−
Circulatory disorders	peripheral vascular disease; unspecified	−	+	−	−	+	+	−
	other chronic ischemic heart disease; unspecified	−	−	+	+	−	−	−
	angina pectoris coronary	−	−	+	+	−	−	−
	atherosclerosis	−	−	+	+	−	−	−
	myocardial infarction	−	−	+	+	−	−	−
	abdominal aortic aneurysm	+	−	−	−	−	−	−
	essential hypertension	−	−	−	−	−	−	+
Reproductive system diseases	uterine leiomyoma	+	−	−	+	+	+	+
	excessive or frequent menstruation	+	−	−	−	−	+	−
Digestive system diseases	diverticulosis	+	−	−	+	+	+	−
	viral hepatitis B	+	−	−	−	+	+	−
Endocrine diseases	hypercholesterolemia	−	−	−	+	−	−	−
	type 2 diabetes	+	+	−	+	+	+	+
Infectious diseases	arthropod-borne diseases	+	−	−	−	−	+	+
Skeletal disease	osteoarthritis; localized	−	+	−	−	+	−	−

Note: "+" means that the trait is positive in the PheWAS analysis of the corresponding SNP, while "-" means negative.

phenoscanner.medschl.cam.ac.uk/) was used to exclude SNPs associated with confounding factors. For Assumption 3, Duplication of the instrumental variables in MR With the aforementioned 7 SNPs in PheWAS will be excluded. It is worth noting that we excluded SNPs associated with confounding factors based on existing research results. If there is any correlation between the selected variants and confounding factors not being reported, then our statistical analysis may have potential limitations.

To conduct MR analysis, the R packages TwoSampleMR (Version 0.5.6) and MRPRESSO (Version 1.0) were utilized as the primary tools. In particular, we employed inverse variance weighting (IVW) and MR Egger regression as the main analysis methods [23]. Under the premise of meeting the three core assumptions, IVW can accurately estimate the causal effects between exposure variable and outcomes. MR Egger can correct for pleiotropy bias, but the estimation results are often biased towards conservatism.

The R-package MR-PRESSO is currently a widely used method for pleiotropy test. Pleiotropy tests were also conducted based on the intercept of MR Egger regression. Additionally, outlier removal was performed based on a leave-one-out exclusion test, correcting the results by removing outliers that surpassed a threshold of $P < 0.05$. Furthermore, Cochran's Q statistic was employed to assess heterogeneity among the results, which helps determine whether variations in the study outcomes are due to differences between the studies or random chance.

By implementing these rigorous analytical approaches and quality control measures, we aimed to derive reliable and unbiased estimates of the causal relationships between smoking and the selected outcome variables by meeting the three core assumptions of MR analysis.

3. Result

3.1. PheWAS analysis

We defined 1,549 binary phenotypes using PheCODE, and 89 traits were removed due to fewer than 20 cases or controls, the results of the seven PheWAS analyses were obtained. A total of 34 phenotypes were significantly positive associated with smoking after Bonferroni correction ($P = 0.05/1460$), and these were clustered in multiple systems, including hematologic, neuropsychiatric, dermatologic, respiratory, circulatory, reproductive, digestive system, endocrine, infectious diseases and skeletal disease (Table 3). Detailed results are shown in the attached table (Supplementary Table S1).

Fifteen of these diseases were significantly associated with at least three of the seven SNPs studied, which were sickle-cell anemia, other hemoglobinopathies, tobacco use disorder, other mental disorder, chronic airway obstruction, obstructive chronic bronchitis, cancer of bronchus; lung, emphysema, pneumococcal pneumonia, peripheral vascular disease; unspecified, uterine leiomyoma, diverticulosis, viral hepatitis B, type 2 diabetes and arthropod-borne diseases. Of these, sickle-cell anemia and type 2 diabetes were significantly associated with six of the seven SNPs of interest. In addition, tobacco use disorder, uterine leiomyoma and three of the respiratory diseases (chronic airway obstruction, obstructive chronic bronchitis and cancer of bronchus; lung) showed significant associations with five out of the seven SNPs. Among the analyzed SNPs, rs16969968, rs13329271, rs6495308, rs1051730, and rs7937 demonstrated 10 or more positive correlation results in the analysis.

The most important findings in this study are that there is a strong correlation between smoking and sickle cell anemia and type 2 diabetes, with sickle-cell anemia results is rs16969968 ($P = 4.45 \times 10^{-58}$), rs12595538 ($P = 1.64 \times 10^{-10}$), rs8032771 ($P = 1.15 \times 10^{-20}$), rs1051730 ($P = 1.34 \times 10^{-42}$), and rs7937 ($P = 6.26 \times 10^{-67}$). In addition, similar to previous findings, there were many significant SNPs between respiratory disease and smoking in this study. These findings enhance our understanding of the wide range of diseases related to smoking and offer meaningful perspectives into possible causal links between cigarette use and different health conditions.

Outcome	Beta	SE	95%CI	P
altered mental status	-6.75×10^{-2}	1.88×10^{-1}		7.20×10^{-1}
arthropod-borne diseases	-2.69×10^{-3}	6.21×10^{-1}		1.00×10^{0}
cancer of bronchus; lung	9.21×10^{-1}	1.25×10^{-1}		2.02×10^{-13}
chronic airway obstruction	8.31×10^{-1}	2.69×10^{-1}		2.00×10^{-3}
diverticulosis	7.03×10^{-2}	5.09×10^{-2}		1.67×10^{-1}
emphysema	1.63×10^{0}	3.73×10^{-1}		1.29×10^{-5}
obstructive chronic bronchitis	4.80×10^{-1}	6.50×10^{-2}		1.62×10^{-13}
peripheral vascular disease; unspecified	1.09×10^{0}	1.92×10^{-1}		1.63×10^{-8}
pneumococcal pneumonia	2.99×10^{-1}	9.48×10^{-2}		1.60×10^{-3}
type 2 diabetes	5.28×10^{-1}	1.88×10^{-1}		5.08×10^{-3}
uterine leiomyoma	-9.62×10^{-2}	6.03×10^{-2}		1.11×10^{-1}
chronic hepatitis	3.06×10^{-1}	3.28×10^{-1}		3.50×10^{-1}

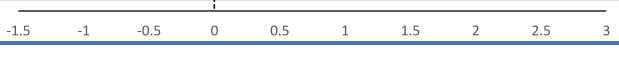


Fig. 2. Mendelian randomization summary results presented in forest plot for smoking and 12 diseases. Estimates were obtained from the inverse-variance weighted method. SE, standard error; CI, confidence intervals.

3.2. Mendelian randomization analysis

To conduct sensitivity analysis, we performed pleiotropy tests and heterogeneity tests, and the results are provided in the attached tables (Supplementary Table S2, Supplementary Table S3). Furthermore, based on the outlier test results from MRPRESSO, we removed the outlier SNP rs73229090 in the analysis of chronic bronchitis and emphysema to mitigate the potential influence of horizontal pleiotropy.

Interestingly, this analysis revealed a risk causality between smoking and diseases (Fig. 2 and Supplementary Table S4) like peripheral vascular disease (IVW: Beta = 1.09, 95% confidence interval (CI) 0.71-1.46, $P = 1.63 \times 10^{-8}$) and type 2 diabetes (IVW: Beta = 0.53, 95% CI 0.16-0.90, $P = 5.08 \times 10^{-3}$), which may not be commonly associated with smoking in the general perception. Based on the results of the IVW analysis in MR, we observed a causal relationship between smoking and several respiratory diseases, including obstructive chronic bronchitis (IVW: Beta = 0.48, 95% CI 0.36-0.61, $P = 1.62 \times 10^{-13}$), cancer of the bronchus; lung (IVW: Beta = 0.92, 95% CI 0.68-1.17, $P = 2.02 \times 10^{-13}$), emphysema (IVW: Beta = 1.63, 95% CI 0.90-2.36, $P = 1.29 \times 10^{-5}$), chronic airway obstruction (IVW: Beta = 0.83, 95% CI 0.30-1.36, $P = 2.00 \times 10^{-3}$) and pneumococcal pneumonia (IVW: Beta = 0.30, 95% CI 0.11-0.49, $P = 1.60 \times 10^{-3}$). This indicates that smoking is indeed a definitive risk factor for these respiratory conditions. However, the findings detected no connection between smoking and psychiatric disorders, reproductive diseases or digestive diseases. Possible reasons require further investigation. Additionally, smoking did not appear to significantly affect the spread of certain non-respiratory infectious diseases within the population.

These findings provide valuable insights into the causal relationship between smoking and various diseases, further emphasizing the detrimental health effects of smoking, particularly on outcomes such as type 2 diabetes and peripheral vascular disease. In addition, we were unable to perform MR for sickle-cell anemia, tobacco use disorder, and other hemoglobinopathies due to data limitations.

4. Discussion

This study aimed to investigate the correlations and potential causative relationships between smoking and various diseases using extensive genotypic and phenotypic data from large population cohorts. Through a multi-SNP integrated PheWAS analysis and MR analyses, we identified associations between smoking and diseases across a broad range of phenotypes. By uncovering 34 phenotypes significantly correlated with smoking through the PheWAS analysis, we shed light on prominent associations such as sickle cell anemia, type 2 diabetes, and respiratory diseases. Expanding on these discoveries, subsequent MR analysis established causal relationships between smoking and 7 diseases.

This study validates the important conclusion that smoking potentially plays a role in the development of type 2 diabetes and peripheral vascular disease. While this study underscores associations between smoking with sickle cell anemia and type 2 diabetes, it acknowledges prior research indicating these connections [24]. Notably, the relationship between smoking and sickle cell anemia remains a subject of debate. While our results and some previous studies support a link [25], contrasting views exist, such as a study found no association [26]. This discrepancy calls for further research to understand the underlying mechanisms and potential interactions. However, it is crucial to address why the majority of PheWAS results were not validated by MR. One possible reason is the presence of horizontal pleiotropy, where genetic variants influence multiple traits through different biological pathways, potentially confounding MR estimates. Additionally, the statistical power to detect causal relationships in MR analysis may be limited for some phenotypes due to smaller sample sizes or weaker genetic instruments. Moreover, the association between smoking with type 2 diabetes and peripheral vascular disease is likely influenced by several factors: firstly, the reactive oxidizing substances present in tobacco smoke can enter the bloodstream, leading to damage to the vascular endothelium. This, in turn, can result in significant activation of leukocytes and platelets, ultimately leading to vascular damage [27,28]. Additionally, nicotine, a key component of tobacco smoke, can directly impact glucose homeostasis and insulin sensitivity. Active smoking has been shown to reduce insulin sensitivity, further exacerbating the risk of developing type 2 diabetes [29]. Furthermore, nicotine influences insulin secretion through its interaction with nicotinic acetylcholine receptors on β-cells, while also promoting apoptosis of pancreatic β-cells [30,21]. These combined effects highlight the potential mechanisms by which smoking may contribute to the development of type 2 diabetes and peripheral vascular disease. The damaging impact of reactive oxidizing substances on the vascular system, coupled with the direct effects of nicotine on glucose regulation and insulin function, underscore the detrimental consequences of smoking on metabolic health.

Indeed, the results point to findings of positive association between smoking and respiratory diseases, such as emphysema, lung cancer, pneumococcal pneumonia, and chronic obstructive pulmonary disease (COPD), which is consistent with existing research [31]. Tobacco smoke has been shown to contain over 4500 harmful substances, including nicotine, tar, carbon dioxide, formaldehyde, and carbon monoxide [27,28]. These substances can have detrimental effects on the respiratory system, leading to various respiratory diseases. The inhalation of tobacco smoke can result in airway constriction, increased production of mucus, damage to the alveolus, and increased susceptibility to lung infections, all of which contribute to the development of respiratory diseases [32]. Furthermore, tobacco smoke triggers the release of reactive oxygen species and reactive nitrogen species from white blood cells, as well as the secretion of pro-inflammatory cytokines, leading to airway inflammation [32]. The mechanisms described above provide insight into how smoking negatively impacts the respiratory system, contributing to the development and progression of respiratory diseases. By confirming these associations and elucidating the underlying molecular and physiological processes, the outcome reinforces the existing understanding of the harmful effects of smoking on respiratory health.

The outcome reveals inconsistencies with previous research regarding the causal relationship between smoking and certain

diseases, such as diverticular disease, mental illness, and chronic hepatitis. It is important to note the limitations and variations in the available studies that have contributed to these inconsistencies. Moreover, it is worth emphasizing that this study had the largest sample size, providing robust evidence for these findings. For instance, a study suggesting a link between smoking and diverticular disease had a small study population, focusing solely on Swedish females [33]. This limited sample size and population specificity may affect the generalizability of the findings. In the case of chronic hepatitis, while one study indicated a causal relationship, residual confounders were not entirely ruled out, potentially affecting the validity of the findings [34]. Similarly, a study summarizing the relationship between smoking and mental illnesses may have overlooked the aspect of causal relationships, leading to incomplete conclusions [35]. Regarding uterine fibroids, existing studies have not found a significant causal relationship with smoking. This aligns with the conclusions, indicating consistency in the current body of evidence [36].

However, it is important to acknowledge the limitations of this study. One significant limitation is the unavailability of suitable data to analyze the causal relationship between smoking and certain conditions, such as sickle cell anemia, tobacco use disorders, and other hemoglobinopathies. This data gap restricts the ability to draw definitive conclusions in these specific areas, and it could potentially affect the validity of results to some extent. Furthermore, the low prevalence of certain diseases within the population resulted in a limited number of affected individuals in the analyzed dataset. This small sample size may affect the statistical power and generalizability of the findings for these particular conditions. Another limitation is the narrow focus on data from European populations, which may limit the generalizability of the conclusions to other ethnic groups or regions. Smoking behaviors and genetic backgrounds vary significantly across different populations, which can influence both exposure levels and susceptibility to smoking-related diseases. For instance, genetic variants linked to smoking behaviors in European populations might not be as influential in Asian or African populations, where different genetic, cultural, and environmental factors may prevail [37,38]. Additionally, the incidence and manifestation of diseases associated with smoking—such as lung cancer, cardiovascular diseases, and diabetes—may differ by ethnicity due to varying genetic predispositions and lifestyle factors [39]. It is crucial, therefore, to conduct similar studies across diverse populations to determine whether the associations observed in this study hold in different ethnic and regional contexts. While we employed a screening process for the PheWAS results by selecting traits that appeared positive at least three times, it is important to note that the remaining traits not meeting this threshold may still have a causal relationship with smoking, and our results may have been incomplete by missing these associations. Thus, these findings should be interpreted with caution, and further investigation and clinical trials are warranted to explore potential associations between smoking and these remaining traits. Acknowledging these limitations is crucial for a comprehensive understanding of the scope and applicability of research findings. Future research endeavors should aim to address these limitations by incorporating larger and more diverse study populations, considering data from various ethnic backgrounds, and conducting focused investigations into specific disease associations that require further exploration.

Overall, we made significant contributions by leveraging a large population-based research cohort to examine the genetic associations between smoking and various diseases. The findings highlight the importance of modifying smoking strategies as an effective approach to control and prevent certain diseases, such as type 2 diabetes, peripheral vascular disease and some respiratory diseases. By elucidating the genetic underpinnings of these associations, we provide valuable insights for developing targeted interventions and public health strategies. This may help provide information for public health debate on smoking policy and prevention strategies to alleviate the burden of smoking-related disease.

Ethics approval and consent to participate

Review and/or approval by an ethics committee was not needed for this study because it solely utilized publicly available genome-wide association study (GWAS) statistics and did not involve any direct human subject involvement or interventions. Therefore, no ethical approval was required.

Consent for publication

Not applicable.

Funding

The Fundamental Research Funds for National Key R&D Program of China (2021YFC2502100), This study was supported by the Joint Institute of smoking and health (2022539200340109), the National Natural Science Foundation of China (82001362), the Hunan Youth Science and Technology Innovation Talent Project (2022RC1070), and the Natural Science Foundation of Hunan Province in China (2021JJ31070).

Data availability statement

This research has been conducted using the UK Biobank Resource under approved application number 103082. After the health-related research is approved by the UK Biobank Access Team, the data and materials are available online at https://www.ukbiobank.ac.uk/. The GWAS dataset used in the research are available in the GWAS Catalog (https://www.ebi.ac.uk/gwas/) and MRC-IEU OpenGWAS (https://gwas.mrcieu.ac.uk/). The smoking GWAS dataset was collected and analyzed by GWAS and Sequencing Consortium of Alcohol and Nicotine use, while the biobank-scale data of diseases GWAS was provided by Longda Jiang et al. [22]. Data

analysis was performed in R version 4.2.2, and the R packages that were mainly used in the study were PheWAS (https://github.com/PheWAS/PheWAS/), TwoSampleMR (https://github.com/MRCIEU/TwoSampleMR), MRPRESSO(https://github.com/rondolab/MR-PRESSO).

CRediT authorship contribution statement

Zixun Lin: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jiayi Xiong:** Writing – review & editing. **Jiaqi Yang:** Data curation. **Yuanfeng Huang:** Supervision. **Jinchen Li:** Resources, Funding acquisition. **Guihu Zhao:** Resources, Funding acquisition. **Bin Li:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful for resources from the High-Performance Computing Centre of Central South University and we would like to thank the International Headache Genetics Consortium for providing us with the migraine GWAS summary data, the GWAS Catlog and the PGC and the MRC-IEU OpenGWAS database for providing other GWAS datasets.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e35649.

References

[1] Office of the Surgeon, G., S. Office on, and Health, Reports of the Surgeon General, in: The Health Consequences of Smoking: A Report of the Surgeon General, Centers for Disease Control and Prevention (US), Atlanta (GA), 2004.
[2] World health organization data, Available from: https://www.who.int/news-room/fact-sheets/detail/tobacco, 2023.
[3] V. Bellou, L. Belbasis, E. Evangelou, Tobacco smoking and risk for pulmonary fibrosis: a prospective cohort study from the UK biobank, Chest 160 (3) (2021) 983–993.
[4] T. Kondo, et al., Effects of tobacco smoking on cardiovascular disease, Circ. J. 83 (10) (2019) 1980–1985.
[5] W. Liu, et al., Secondhand smoking and neurological disease: a meta-analysis of cohort studies, Rev. Environ. Health 36 (2) (2021) 271–277.
[6] R. Piipari, et al., Smoking and asthma in adults, Eur. Respir. J. 24 (5) (2004) 734–739.
[7] A. Hjern, et al., Does tobacco smoke prevent atopic disorders? A study of two generations of Swedish residents, Clin. Exp. Allergy 31 (6) (2001) 908–914.
[8] J. Ham, et al., Cigarette smoke aggravates asthma by inducing memory-like type 3 innate lymphoid cells, Nat. Commun. 13 (1) (2022) 3852.
[9] B. Mappin-Kasirer, et al., Tobacco smoking and the risk of Parkinson disease: a 65-year follow-up of 30,000 male British doctors, Neurology 94 (20) (2020) e2132–e2138.
[10] B. Zhou, et al., Cigarette smoking and the risk of endometrial cancer: a meta-analysis, Am. J. Med. 121 (6) (2008) 501–508.e3.
[11] L. Bastarache, J.C. Denny, D.M. Roden, Phenome-wide association studies, JAMA 327 (1) (2022) 75–76.
[12] J. Bowden, M.V. Holmes, Meta-analysis and Mendelian randomization: a review, Res. Synth. Methods 10 (4) (2019) 486–496.
[13] R. Collins, What makes UK Biobank special? Lancet 379 (9822) (2012) 1173–1174.
[14] S. Yuan, et al., Health effects of milk consumption: phenome-wide Mendelian randomization study, BMC Med. 20 (1) (2022) 455.
[15] W.Q. Wei, et al., Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record, PLoS One 12 (7) (2017) e0175508.
[16] L.J. Bierut, et al., Variants in nicotinic receptors and risk for nicotine dependence, Am. J. Psychiatr. 165 (9) (2008) 1163–1171.
[17] N. Matoba, et al., GWAS of smoking behaviour in 165,436 Japanese people reveals seven new loci and shared genetic architecture, Nat Hum Behav 3 (5) (2019) 471–477.
[18] R.J. Hung, et al., A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25, Nature 452 (7187) (2008) 633–637.
[19] D.W. Belsky, et al., Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study, JAMA Psychiatr. 70 (5) (2013) 534–542.
[20] UK Biobank data. Available from:: https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/genetic-data.
[21] E. Somm, Nicotinic cholinergic signaling in adipose tissue and pancreatic islets biology: revisited function and therapeutic perspectives, Arch. Immunol. Ther. Exp. 62 (2) (2014) 87–101.
[22] L. Jiang, et al., A generalized linear mixed model association tool for biobank-scale data, Nat. Genet. 53 (11) (2021) 1616–1621.
[23] C.H. Lee, et al., Comparison of two meta-analysis methods: inverse-variance-weighted average and weighted sum of Z-scores, Genomics Inform 14 (4) (2016) 173–180.
[24] X.T. Xie, et al., Impact of cigarette smoking in type 2 diabetes development, Acta Pharmacol. Sin. 30 (6) (2009) 784–787.
[25] R.C. Young Jr., et al., Smoking is a factor in causing acute chest syndrome in sickle cell anemia, J. Natl. Med. Assoc. 84 (3) (1992) 267–271.
[26] O. Castro, et al., The acute chest syndrome in sickle cell disease: incidence and risk factors. The Cooperative Study of Sickle Cell Disease, Blood 84 (2) (1994) 643–649.
[27] Y. Li, S.S. Hecht, Carcinogenic components of tobacco and tobacco smoke: a 2022 update, Food Chem. Toxicol. 165 (2022) 113179.
[28] N. Katsiki, et al., Smoking and vascular risk: are all forms of smoking harmful to all types of vascular disease? Publ. Health 127 (5) (2013) 435–441.
[29] S. Mukharjee, S. Bank, S. Maiti, Chronic tobacco exposure by smoking develops insulin resistance, Endocr., Metab. Immune Disord.: Drug Targets 20 (6) (2020) 869–877.
[30] J. Maddatu, E. Anderson-Baucum, C. Evans-Molina, Smoking and the risk of type 2 diabetes, Transl. Res. 184 (2017) 101–107.

[31] L. Jayes, et al., SmokeHaz: systematic reviews and meta-analyses of the effects of smoking on respiratory health, Chest 150 (1) (2016) 164–179.
[32] F. Soleimani, et al., Content of toxic components of cigarette, cigarette smoke vs cigarette butts: a comprehensive systematic review, Sci. Total Environ. 813 (2022) 152667.
[33] F. Hjern, A. Wolk, N. Håkansson, Smoking and the risk of diverticular disease in women, Br. J. Surg. 98 (7) (2011) 997–1002.
[34] E. Tsochatzis, et al., Smoking is associated with steatosis and severe fibrosis in chronic hepatitis C but not B, Scand. J. Gastroenterol. 44 (6) (2009) 752–759.
[35] A. Minichino, et al., Smoking behaviour and mental health disorders–mutual influences and implications for therapy, Int. J. Environ. Res. Publ. Health 10 (10) (2013) 4790–4811.
[36] F. Chiaffarino, et al., Cigarette smoking and risk of uterine myoma: systematic review and meta-analysis, Eur. J. Obstet. Gynecol. Reprod. Biol. 197 (2016) 63–71.
[37] G.R.B. Saunders, et al., Genetic diversity fuels gene discovery for tobacco and alcohol use, Nature 612 (7941) (2022) 720–724.
[38] L.M. Evans, et al., Genetic architecture of four smoking behaviors using partitioned SNP heritability, Addiction 116 (9) (2021) 2498–2508.
[39] Z. Zhu, et al., Determination of genetic correlation between tobacco smoking and coronary artery disease, Front. Psychiatr. 14 (2023) 1279962.