Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Transcriptome-Proteome Decoupling in Smoking-Induced Cardiovascular Disease: A Multi-Omics Mendelian Randomization Study

Shuo Ying
Tianjin Medical University
Qiang Chen
Tianjin Chest Hospital
Xiaogang Liu
Tianjin Chest Hospital
Peng Zhang
Tianjin Chest Hospital
Naikuan Fu
Tianjin Chest Hospital
Yingwu Liu
liuyingwu1203@163.com

Tianjin Third Central Hospital

Research Article

Keywords:

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11475023/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Title Page

Title: Transcriptome-Proteome Decoupling in Smoking-Induced Cardiovascular Disease: A Multi-Omics Mendelian Randomization Study

Running title: Multi-Omics Decoupling of Smoking and CVD

Authors and institutional affiliations: Shuo Ying M.D.[1,2*], Qiang Chen M.D.[3*], Xiaogang Liu M.D.[2], Peng Zhang M.D.[2&], Naikuan Fu M.D.[2&], Yingwu Liu M.D.[1&]

1. The Third Central Clinical College of Tianjin Medical University, Tianjin, 300170, China.

2. Department of Cardiology, Tianjin Chest Hospital, Affiliated Chest Hospital of Tianjin University, Tianjin, 300222, China.

3. Department of Respiratory and Critical Medicine, Tianjin Chest Hospital, Affiliated Chest Hospital of Tianjin University, Tianjin, 300222, China.

Corresponding authors:

Peng Zhang, No.261 Taierzhuang South Road, Jinnan District, Tianjin, 300222, China. Email: pzhang0822@163.com

Naikuan Fu, No.261 Taierzhuang South Road, Jinnan District, Tianjin, 300222, China. Email: nkf0822@163.com

Yingwu Liu, No.83 Jintang Road, Hedong District, Tianjin, 300170, China. Email: liuyingwu1203@163.com

[*]These authors have contributed equally to this work and share first authorship.

[&]These authors contributed equally to this work and should be considered co-corresponding authors.

Institutional email address: sxkyykjk@tj.gov.cn.

Transcriptome-Proteome Decoupling in Smoking-Induced Cardiovascular Disease: A Multi-Omics Mendelian Randomization Study

Abstract

Objective: Smoking is a major modifiable cardiovascular risk factor. We aimed to dissect the causal relationships between multiple smoking phenotypes and cardiovascular diseases (CVDs), and critically evaluate whether smoking-induced molecular alterations propagate linearly from DNA methylation to gene expression and circulating proteins.

Methods: We performed bidirectional two-sample Mendelian randomization (MR) for 11 smoking behaviors and 12 CVD outcomes, and applied multivariable MR (MVMR) to estimate direct effects independent of cardiometabolic confounders. To investigate molecular mechanisms, we integrated summary-data-based MR (SMR) with blood cis-expression quantitative trait locus (eQTL) data and performed Bayesian colocalization between smoking-associated DNA methylation (mQTL), gene expression, and plasma/cerebrospinal fluid (CSF) protein QTL (pQTL). Finally, we systematically interrogated the cross-platform coverage of candidate proteins across major proteomic datasets.

Results: Genetically proxied "ever smoked" and "cigarettes per day" significantly increased the risk of CAD, PAD, and HF, with direct effects persisting independent of cardiometabolic confounders. eQTL-SMR identified 640 genes shared across CVD subtypes. However, this cascade was not linear. While 11 CpG-gene pairs showed robust mQTL-eQTL colocalization, the classical smoking marker AHRR showed a complete absence of colocalization (PP.H3 = 0.94–1.00), indicating that epigenetic imprints reflect exposure rather than causal mediators. A second layer of systemic decoupling was observed between transcription and the proteome. Among the seven candidate genes with robust mQTL-eQTL colocalization, six were entirely absent from the SomaLogic 7k official menu. For the sole covered gene, SPTLC2, eQTL-pQTL colocalization was negative in both plasma (PP.H1 = 0.65) and CSF (minimum P = 4.96×10^{-4}). In parallel, pQTL MR independently identified 18 subtype-specific causal

proteins (e.g., ABO for DVT/PE, CTF1/VPS29 for CAD) that were not direct downstream effectors of smoking.

Conclusion: Smoking behaviours exert heterogeneous effects on CVDs through shared upstream transcriptional programs with subtype-specific downstream protein effectors. However, this process is characterized by multi-layer decoupling—driven by the systematic exclusion of intracellular regulatory proteins from current proteomic platforms. Our findings redefine the molecular architecture of smoking-induced CVD and highlight the necessity of distinguishing molecular enrichment from shared causal continuity in multi-omic MR studies.

Key words: Smoking, cardiovascular disease, Mendelian randomization, multi-layer decoupling, eQTL-SMR, DNA methylation, pQTL MR

Introduction

Cardiovascular disease (CVD) remains the leading cause of global mortality. Established major risk factors include type 2 diabetes mellitus (T2DM), overweight/obesity, hypertension, dyslipidemia, and smoking [1]. Tobacco use alone accounts for over seven million annual deaths and reduces life expectancy by at least a decade [2, 3].

Smoking constitutes a major modifiable risk factor for the initiation and progression of CVDs [4]. Its impact on the cardiovascular system is pleiotropic, elevating the risk of atrial fibrillation (AF), coronary artery disease (CAD), hypertension, peripheral artery disease (PAD), heart failure (HF), sudden cardiac death (CD), aortic aneurysm (AA), and stroke by 2- to 4-fold compared to non-smokers [5]. However, several fundamental questions persist. First, most evidence comes from observational studies, which are susceptible to confounding and reverse causation. Second, ‘smoking’ is heterogeneous—initiation, intensity, second-hand exposure, and product type may have differential effects on specific CVD subtypes. Third, the pathways through which smoking operates are pleiotropic; it may act directly on the vasculature or indirectly by promoting T2DM, hypertension, body mass index (BMI) and dyslipidemia. Whether these metabolic factors fully explain the smoking–CVD

association, or whether additional molecular mechanisms are involved, remains unclear.

Mendelian randomization (MR) uses genetic variants as instrumental variables to strengthen causal inference under the assumption of no pleiotropy [6]. Previous MR studies have examined smoking and broad CVD outcomes, but a comprehensive dissection that simultaneously addresses behavioural heterogeneity, a wide spectrum of CVDs, mediation by metabolic traits, and potential molecular mediators is lacking.

Here we present an integrative multi-omics MR study with four objectives: (1) to estimate causal effects of 11 smoking phenotypes on 12 CVDs using bidirectional two-sample MR; (2) to estimate direct effects independent of seven cardiometabolic confounders using MVMR among FDR-surviving exposure–outcome pairs; (3) to identify shared transcriptional programs using eQTL-SMR; and (4) to identify subtype-specific causal proteins using pQTL MR and Bayesian colocalization. This framework moves beyond simple causal inference to map the multi-layered molecular architecture linking smoking to cardiovascular morbidity. Additionally, we aim to critically evaluate the continuity of the "DNA methylation → gene expression → protein → CVD" axis, testing whether the genetic architectures driving these multi-omic layers are linearly linked or decoupled.

Methods

Study Design

We performed comprehensive two-sample MR analyses using publicly available summary-level genome-wide association studies (GWAS) data. Ethical approval was not required. The MR design relies on three core assumptions (Figure 1): (1) genetic instruments are robustly associated with the exposure (smoking behaviors); (2) instruments are independent of potential confounders; and (3) instruments affect the outcome only through the exposure (no horizontal pleiotropy) [7]. All statistical analyses were conducted using R software (version 4.5.0).

Fig. 1 Study design. Genetic instruments for 11 smoking behaviours were used to estimate causal effects on 12 cardiovascular diseases using two-sample MR. MVMR

estimated direct effects independent of seven cardiometabolic confounders. Molecular mechanisms were investigated using eQTL-based SMR, pQTL MR, and Bayesian colocalization. Three core MR assumptions are indicated: (1) instruments are strongly associated with the exposure; (2) instruments are independent of confounders; (3) instruments affect the outcome only through the exposure (no horizontal pleiotropy). Red X marks the two core assumptions that are required but not directly tested. Black arrows denote causal effects; red arrows denote MVMR adjustment; purple arrows denote molecular mechanism investigation. Box colours indicate exposure (blue), outcome (green), confounder (yellow), and molecular layer (purple). SNP = single nucleotide polymorphism; MVMR = multivariable Mendelian randomization; SMR = summary-data-based Mendelian randomization; eQTL = expression quantitative trait locus; pQTL = protein quantitative trait locus; PP.H4 = posterior probability of a shared causal variant; GSCAN = GWAS and Sequencing Consortium of Alcohol and Nicotine use; UKB = UK Biobank; T2DM = type 2 diabetes mellitus; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; TG = triglycerides; BP = blood pressure; BMI = body mass index; Lp(a) = lipoprotein(a); DIAMANTE = DIAMANTE Consortium; GLGC = Global Lipids Genetics Consortium; ICBP = International Consortium for Blood Pressure; GIANT = Genetic Investigation of ANthropometric Traits; INTERVAL = INTERVAL SomaScan study; CHRNA3/CHRNA5 = cholinergic receptor nicotinic alpha 3 / alpha 5 subunits.

Data Sources

Smoking behaviors were categorized based on history, intensity, and exposure context. Exposures included: 'ever smoked', secondhand smoke exposure (' cigarette smoke from other people smoking: sometime and often'), smoking intensity (' cigarettes per day', 'smokes on most or all days', 'light smokers' defined as ≥100 cigarettes lifetime), location of exposure (' exposure to tobacco in household and outside home'), and tobacco type ('hand-rolled cigarettes', 'manufactured cigarettes', 'cigars or pipes'). CVD outcomes comprised AF, CAD, deep venous thrombosis (DVT) / pulmonary embolism (PE), PAD, HF, CD, AA, stroke, ventricular tachycardia (VA), tachycardia, bradycardia

and dilated cardiomyopathy (DCM).

GWAS summary statistics were obtained from the MRC-IEU OpenGWAS project, the GWAS Catalog, the GSCAN Consortium, the FinnGen Consortium, DIAMANTE/DIAGRAM, GLGC, ICBP, and GIANT. Blood cis-eQTL data were obtained from the eQTLGen Consortium (N = 31,684). Plasma cis-pQTL data were derived from the INTERVAL SomaScan dataset (Sun et al. 2018; N = 3,301), accessed via the IEU OpenGWAS platform (https://gwas.mrcieu.ac.uk/). The 1000 Genomes Project European reference panel was used for LD clumping and colocalization. Detailed dataset characteristics and the specific source for each trait are provided in Table 1. All participants were of European descent.

Instrumental Variables (IVs) Selection

For each smoking behaviour, independent genetic instruments were selected at $P < 5\times10^{-8}$ and clumped at $r^2 < 0.001$ within 10,000 kb. Instrument strength was quantified by the F-statistic; SNPs with $F < 10$ were excluded. Palindromic SNPs were removed during harmonisation.

Assessment of core MR assumptions: (1) relevance was ensured by the genome-wide significance threshold and the F-statistic criterion; (2) independence from confounders was systematically tested by querying each selected SNP against five established behavioural/social confounders (educational attainment, household income, alcohol consumption, any mental disorder, physical activity); (3) exclusion restriction was evaluated through multiple sensitivity analyses (MR-Egger intercept, MR-PRESSO and leave-one-out).

Screening of instrumental SNPs against behavioural and social confounders

Each instrumental SNP for the three primary smoking exposures ("ever smoked", "cigarettes per day", and "exposure to tobacco smoke outside home") was queried against five behavioural/social confounders: educational attainment, average total household income before tax, alcohol consumption, any mental disorder, and physical activity. Multiple testing was controlled by the Benjamini–Hochberg FDR within each exposure–confounder pair. SNPs with FDR < 0.05 for at least one confounder were

excluded when sufficient instruments remained; SNPs significant at the nominal level only were retained but flagged.

Statistical Analysis

Univariable MR (UVMR)

Analyses were primarily performed using the inverse-variance weighted (IVW) method under a random-effects model, which provides the most precise estimate when all instruments are valid [8]. To ensure robustness, we supplemented this with the weighted median (WM) estimator, which yields consistent causal estimates even if up to 50% of the genetic instruments are invalid [9], and MR-Egger regression, which provides a test for directional pleiotropy via its intercept and a causal estimate valid under the InSIDE assumption [10]. Effect sizes are presented as odds ratios (ORs) with 95% confidence intervals (CIs).

Sensitivity analyses

Heterogeneity was assessed using Cochran's Q statistic. Horizontal pleiotropy was evaluated using the MR-Egger intercept test. MR-PRESSO was used to detect and correct for outliers; both raw and outlier-corrected estimates were reported when outliers were identified (Supplementary Table S2). Directionality was evaluated using Steiger filtering (Supplementary Table S3). Leave-one-out and single-SNP analyses were used to identify influential variants (Supplementary Table S4). These sensitivity analyses were performed for the primary analyses where sufficient instruments were available.

Multiple testing correction

To control for false discovery due to multiple comparisons, we applied the Benjamini-Hochberg false discovery rate (FDR) correction within each smoking behaviour–CVD outcome matrix. Associations with FDR $q < 0.05$ were considered robust, while those with $q > 0.05$ but nominal $p < 0.05$ were interpreted as suggestive and discussed with caution.

Multivariable MR (MVMR)

To estimate direct effects independent of candidate cardiometabolic mediators, we

performed MVMR using the IVW method with robust standard errors. Conditional ORs were reported as estimates of direct effects. Single-confounder models adjusted for T2DM, hypertension, BMI, HDL-C, LDL-C, TG, or Lp(a). Combined models included a metabolic model (T2DM + BMI), a lipid model (LDL-C + TG + HDL-C), and an all-available model (T2DM + BMI + hypertension + LDL-C + TG). Conditional F-statistics were used to assess instrument strength; models with conditional $F < 10$ were flagged as exploratory. Sample overlap was assessed using overlap proportion and bias metrics.

Integration with expression quantitative trait locus (eQTL) data

For eQTL- summary-data-based Mendelian randomization (SMR), blood cis-eQTL data from the eQTLGen consortium (N = 31,684) were integrated with GWAS summary statistics for five CVD subtypes (CAD, DCM, DVT/PE, PAD, and HF). SNPs with cis-eQTL $P < 5 \times 10^{-8}$ were selected as instruments and LD-pruned ($r^2 < 0.9$) using the 1000 Genomes European reference panel. Genes with SMR $P < 0.05$ were considered significant. HLA region genes (chr6:25–35 Mb) were excluded from the shared-gene and enrichment analyses. KEGG and GO biological process enrichment were performed using clusterProfiler. Multiple testing was controlled by the Benjamini–Hochberg procedure; adjusted $p < 0.05$ was considered significant. The q value was also reported in Supplementary Tables but was not used for primary significance calling.

DNA methylation and colocalization analysis

Smoking-associated CpGs were obtained from Joehanes et al. (2016), including 18,760 CpGs associated with current versus never smoking at FDR < 0.05. CpGs within ±100 kb of the 640 shared eQTL-SMR genes were selected as candidates. Blood cis-mQTL data were queried from the GoDMC consortium ($N \approx 27,750$). Blood cis-eQTL data were extracted from eQTLGen (N = 31,684). For each CpG-gene pair, common SNPs between mQTL and eQTL datasets were aligned by effect allele, and Bayesian colocalization was performed using coloc.abf. PP.H4 > 0.75 was considered strong evidence for a shared causal variant, and PP.H4 > 0.5 was considered suggestive.

Sensitivity analysis was conducted by removing the lead mQTL and eQTL SNPs and re-running colocalization; pairs retaining PP.H4 > 0.75 were classified as multi-SNP supported (identifying the core seven candidate genes). For genes with suggestive colocalization in the primary analysis, the nearest three CpGs were additionally tested. To dissect the allelic heterogeneity at the classical smoking marker locus AHRR (where robust colocalization was absent), we independently extracted the cis-mQTL lead SNPs from GoDMC and queried their corresponding cis-eQTL effects in eQTLGen. Wald ratio estimates were calculated for these independent mQTL lead SNPs to formally evaluate their causal effect on AHRR expression, testing the hypothesis that the mQTL and eQTL signals were driven by distinct genetic architectures.

protein quantitative trait locus (pQTL) MR

To investigate whether circulating protein levels mediate the observed smoking–CVD associations, we performed two-sample MR using cis-pQTL instruments from the INTERVAL SomaScan dataset. Proteins with at least three cis-pQTL instruments at $P < 5\times10^{-8}$ were analysed. IVW was used as the primary method, with weighted median and MR-Egger as sensitivity analyses. FDR was controlled within each outcome.

Cross-platform coverage and tissue-specific pQTL analysis

To investigate whether the absence of eQTL-pQTL signals reflected biological or technical limitations, we systematically interrogated the cross-platform coverage of the seven candidate genes across the SomaLogic v4.1 (7k) official menu and multiple independent cohorts (UKB-PPP Olink, STARNET, and Fenland). To test for tissue-specific genetic regulation, we extracted cis-pQTL summary statistics for SPTLC2 (Analyte X7267.2, SeqId 7737-76) from the cerebrospinal fluid (CSF) proteomic dataset (NIAGADS NG00130, N=3,506). Additionally, we performed Bayesian colocalization between eQTLGen SPTLC2 cis-eQTL and INTERVAL plasma cis-pQTL data to formally test for shared causal variants.

Smoking-to-protein MR

To determine whether the CVD-causal proteins identified by pQTL MR act as downstream effectors of smoking exposure, we performed two-sample MR using the

four smoking behaviours as exposures and the proteins that reached FDR significance in the pQTL MR analysis as outcomes. The smoking instruments were the same confounder-filtered SNPs used in the primary UVMR analyses. pQTL summary statistics were obtained from the INTERVAL SomaScan dataset via the IEU OpenGWAS platform. The IVW method was used as the primary analysis, with weighted median and MR-Egger as sensitivity analyses. Heterogeneity was assessed using Cochran's Q statistic, and horizontal pleiotropy was evaluated using the MR-Egger intercept. Multiple testing was controlled by the Benjamini–Hochberg FDR within each smoking behaviour.

Functional enrichment analysis

GO biological process and KEGG pathway enrichment were performed using the clusterProfiler package for the union SMR genes, the shared SMR genes, and the causal proteins identified by pQTL MR. Reactome pathway enrichment was additionally performed for the pQTL proteins using ReactomePA. Multiple testing was controlled using the Benjamini-Hochberg FDR.

Bayesian colocalization

Bayesian colocalization was performed at the CHRNA3/CHRNA5 locus (chr15:78.8 Mb) using the coloc package. We extracted SNPs within ±200 kb of the lead SNP (rs9788721) from both exposure and outcome GWAS, aligned effect alleles, and computed posterior probabilities for five hypotheses: H0 (no association), H1 (association with smoking only), H2 (association with CVD only), H3 (two independent causal variants), and H4 (shared causal variant). A PP.H4 > 0.75 was considered strong evidence for a shared causal variant; PP.H4 > 0.5 was considered suggestive.

Results

Characteristics of Genetic Instruments

Summary data for all exposures and outcomes were successfully retrieved (Table 1).

Table 1 Baseline Characteristics of the Instrumental Variables and GWAS datasets.

Trait	Sample size (n)	Case (n)	Control (n)	SNP (n)	Ethnicity	Gender	Data source

Ever smoked	461,066	280,508	180,558	9,851,867	European	Males and Females	Ben Elsworth
Workplace had a lot of cigarette smoke from other people smoking: Often	89,803	14,941	74,862	13,566,864	European	Males and Females	Neale lab
Workplace had a lot of cigarette smoke from other people smoking: Sometimes	90,168	45,053	45,115	13,566,864	European	Males and Females	Neale lab
Cigarettes per day	143,210	NA	NA	13,833,519	European	Males and Females	GSCAN
Tobacco smoking: Smokes on most or all days	91,353	2,124	89,229	10,552,527	European	Males and Females	Neale lab
Light smokers, at least 100 smokes in lifetime	123,894	57,021	66,873	9,851,867	European	Males and Females	Ben Elsworth
Exposure to tobacco smoke outside home	391,502	NA	NA	9,851,867	European	Males and Females	Ben Elsworth
Smoking/smokers in household	425,516	NA	NA	9,851,867	European	Males and Females	Ben Elsworth
Type of tobacco currently smoked:	27,583	6,469	21,114	12,492,318	European	Males and Females	Neale lab

Hand-rolled cigarettes							
Type of tobacco currently smoked: Manufactured cigarettes	27,583	19,103	8,480	13,529,779	European	Males and Females	Neale lab
Type of tobacco currently smoked: Cigars or pipes	27,583	2,011	25,572	10,458,916	European	Males and Females	Neale lab
Atrial fibrillation	1,030,836	60,620	970,216	33,519,037	European	NA	Nielsen JB
Coronary artery disease (SPA correction)	352,063	NA	NA	11,027,841	European	NA	Mbatchou J
DVT of lower extremities and pulmonary embolism	361,194	4,319	356,875	11,783,033	European	Males and Females	Neale lab
Peripheral artery disease	483,078	7,114	475,964	24,186,090	European	NA	Sakaue S
Heart failure	977,323	47,309	930,014	7,773,021	European	NA	Shah S
Death due to cardiac causes	218,792	7,563	211,229	16,380,466	European	Males and Females	NA
Aortic aneurysm	218,792	2,825	215,967	16,380,466	European	Males and Females	NA
Stroke	446,696	40,585	406,111	8,211,693	European	NA	Malik R
Ventricular arrhythmia	328,216	1,018	327,198	19,076,016	European	NA	Sakaue S
Diagnoses - secondary ICD10: R00.0	463,010	1,005	462,005	9,851,867	European	Males and Femal	Ben Elsworth

Tachycardia, unspecified						es	
Diagnoses - secondary ICD10: R00.1 Bradycardia, unspecified	463,010	1,254	461,756	9,851,867	European	Males and Females	Ben Elsworth
Dilated cardiomyopathy	355,381	1,444	353,937	19,080,278	European	NA	Sakaue S
Average total household income before tax	397,751	NA	NA	9,851,867	European	Males and Females	Ben Elsworth
Years of schooling	766,345	NA	NA	10,101,242	European	Males and Females	Lee
Any mental disorder	361,194	4,304	356,890	11,777,016	European	Males and Females	Neale lab
Alcohol	64,979	NA	NA	9,851,867	European	Males and Females	Ben Elsworth
Physical activity	78,007	NA	NA	7,987,378	European	Males and Females	Howe LJ
body mass index	681,275	NA	NA	2,336,260	European	Males and Females	Yengo, L
systolic blood pressure	757,601	NA	NA	7,088,083	European	Males and Females	Evangelou, E
Lipoprotein (a) levels	348,806	NA	NA	4,199,443	European	NA	Barton AR
T2DM	455,297	54,998	400,299	21,476,515	European	Males and Females	Mahajan A

TG	864,240	NA	NA	34,673,394	European	Males and Females	GLGC
LDL-C	842,660	NA	NA	34,553,199	European	Males and Females	GLGC
HDL-C	888,227	NA	NA	35,200,665	European	Males and Females	GLGC

isms; T2DM: type 2 diabetes mellitus; TG: triglycerides; LDL-C: low-density lipoprotein cholesterol; HDL-C: high-density lipoprotein cholesterol

For the primary smoking behaviours, we identified sufficient independent genetic instruments (F-statistic > 10) for MR analyses. After clumping and harmonisation, all instruments had F-statistics > 10, indicating robust instrument strength and negligible weak instrument bias. Screening of all instrumental SNPs against five behavioural/social confounders (educational attainment, household income, alcohol consumption, any mental disorder, and physical activity) identified FDR-significant associations in Ever smoked (28/84 SNPs), Exposure to tobacco smoke outside home (4/7 SNPs), and Cigarettes per day (1/9 SNPs), but not in Light smokers (0/3 SNPs). Confounder-associated SNPs were excluded where instrument availability permitted (Ever smoked: 28 excluded; Exposure to tobacco smoke outside home: 4 excluded; Cigarettes per day: 1 excluded; Light smokers: 0 excluded). All retained instruments had F-statistics > 10. The complete SNP-by-confounder screening results, including FDR q-values and keep/exclude decisions, are provided in Supplementary Table S1.

UVMR analyses

After FDR correction across the 35 estimable exposure–outcome tests (Benjamini–Hochberg; Bonferroni threshold = 0.05/35 = 1.43×10^{-3}), genetically proxied "ever smoked" remained significantly associated with three cardiovascular outcomes: CAD (IVW OR = 2.84, 95% CI 1.50–5.38, P = 1.37×10^{-3}, q = 0.02; Bonferroni-significant), PAD (OR = 2.94, 95% CI 1.46–5.94, P = 2.60×10^{-3}, q = 0.02), and HF (OR = 2.17, 95% CI 1.33–3.55, P = 2.02×10^{-3}, q = 0.02). DVT/PE showed a

nominally significant association that did not survive FDR correction (OR = 1.01, 95% CI 1.00–1.03, P = 1.46×10^{-2}, q = 0.06). No significant associations were observed for AF (OR = 1.27, 95% CI 0.84–1.92, P = 0.26, q = 0.51), CD (OR = 2.12, 95% CI 0.95–4.75, P = 0.07, q = 0.23), AA (OR = 2.92, 95% CI 0.84–10.16, P = 0.09, q = 0.23), stroke (OR = 1.55, 95% CI 0.87–2.76, P = 0.14, q = 0.30), VA (OR = 0.79, 95% CI 0.19–3.25, P = 0.75, q = 0.84), tachycardia (OR = 1.00, 95% CI 1.00–1.01, P = 0.38, q = 0.62), bradycardia (OR = 0.99, 95% CI 0.99–1.00, P = 0.08, q = 0.23), or DCM (OR = 5.48, 95% CI 0.82–36.56, P = 0.08, q = 0.23).

For "cigarettes per day", PAD (OR = 2.77, 95% CI 1.73–4.44, P = 2.42×10^{-5}, q = 8.48×10^{-4}) and HF (OR = 1.35, 95% CI 1.17–1.56, P = 5.24×10^{-5}, q = 9.17×10^{-4}) survived both FDR and Bonferroni correction. No significant associations were observed for AF (OR = 1.00, 95% CI 0.87–1.14, P = 0.98, q = 0.98), CAD (OR = 1.19, 95% CI 0.95–1.49, P = 0.13, q = 0.30), DVT/PE (OR = 1.00, 95% CI 1.00–1.01, P = 0.67, q = 0.81), CD (OR = 0.92, 95% CI 0.66–1.27, P = 0.61, q = 0.77), AA (OR = 1.59, 95% CI 0.79–3.19, P = 0.19, q = 0.40), stroke (OR = 1.10, 95% CI 0.88–1.37, P = 0.43, q = 0.65), VA (OR = 0.83, 95% CI 0.49–1.41, P = 0.50, q = 0.67), bradycardia (OR = 1.00, 95% CI 1.00–1.00, P = 0.54, q = 0.71), or DCM (OR = 1.70, 95% CI 0.93–3.09, P = 0.09, q = 0.23). The association between "cigarettes per day" and tachycardia could not be estimated because only two instruments remained after harmonisation.

For "exposure to tobacco smoke outside home", only 2–3 instruments remained after behavioural/social confounder screening. IVW estimates reached FDR significance for PAD (OR = 28.29, 95% CI 2.68–298.32, P = 5.42×10^{-3}, q = 0.03) and stroke (OR = 0.15, 95% CI 0.04–0.62, P = 8.66×10^{-3}, q = 0.04), and were nominally significant for CAD (OR = 8.61, 95% CI 1.47–50.64, P = 0.02, q = 0.07) .No significant associations were observed for AF (OR = 2.67, 95% CI 0.33–21.29, P = 0.35, q = 0.62), DVT/PE (OR = 0.99, 95% CI 0.96–1.03, P = 0.77, q = 0.84), HF (OR = 0.52, 95% CI 0.08–3.14, P = 0.47, q = 0.67), CD (OR = 1.57, 95% CI 0.09–26.67, P = 0.75, q = 0.84), AA (OR = 0.16, 95% CI 0.00–11.58, P = 0.39, q = 0.62), VA (OR = 5.53, 95% CI 0.04–767.65, P = 0.50, q = 0.67), tachycardia (OR = 1.00, 95% CI 0.98–1.02, P = 0.89, q =

0.91), bradycardia (OR = 0.99, 95% CI 0.96–1.01, P = 0.32, q = 0.59), or DCM (OR = 2.30, 95% CI 0.01–1021.89, P = 0.79, q = 0.84). (Table 2).

Table 2 Univariable MR associations between smoking behaviours and cardiovascular diseases

Smoking type		Ever smoked			Cigarettes smoked per day			Exposure to tobacco smoke outside home		
Methods		IVW	WM	MR-Egger	IVW	WM	MR-Egger	IVW	WM	MR-Egger
Outcomes	SNPs (n)	53			8			3		
AF	OR	1.27	1.21	0.94	1.00	0.95	0.82	2.67	1.84	1.75
	95% CI	0.84-1.92	0.78-1.89	0.06-13.86	0.87-1.14	0.82-1.09	0.64-1.05	0.33-21.29	0.42-8.07	0.00-13333.37
	P	0.26 (0.51)	0.39	0.97	0.98(0.98)	0.45	0.16	0.35(0.62)	0.42	0.92
CAD	SNPs (n)	46			8			2		
	OR	2.84	2.28	14.75	1.19	1.12	0.86	8.61	-	-
	95% CI	1.50-5.38	1.17-4.46	0.20-1084.52	0.95-1.49	0.93-1.35	0.57-1.32	1.47-50.64	-	-
	P	1.37×10^{-3}	0.02	0.23	0.13(0.30)	0.22	0.52	0.02(0.07)	-	-

		(0.02)**								
	SNPs (n)	53			8			3		
DVT/PE	OR	1.01	1.01	1.04	1.00	1.00	1.00	0.99	0.99	0.99
	95% CI	1.00-1.03	0.99-1.02	0.96-1.12	1.00-1.01	0.99-1.01	0.99-1.01	0.96-1.03	0.95-1.04	0.87-1.12
	P	1.46×10^{-2}(0.06)	0.31	0.32	0.67(0.81)	0.87	0.43	0.77(0.84)	0.73	0.90
	SNPs (n)	53			8			3		
PAD	OR	2.94	2.26	0.02	2.77	2.01	1.71	28.29	30.80	5.90
	95% CI	1.46-5.94	0.86-5.98	0.00-1.82	1.73-4.44	1.46-2.77	0.66-4.46	2.68-298.32	1.72-550.10	0.002-22723.61
	P	2.60×10^{-3}(0.02)*	0.099	0.097	2.42×10^{-5}(8.48×10^{-4})**	2.11×10^{-5}	0.31	5.42×10^{-3}(0.03)*	0.02	0.75
	SNPs (n)	52			8			3		
VA	OR	0.79	0.65	> 100	0.83	0.85	0.68	5.53	2.75	0.05

	95% CI	0.19-3.25	0.09-4.45	0.27-1086 3517.08	0.49-1.41	0.47-1.55	0.23-2.00	0.04-767.65	0.01-702.79	0.00-6254 9983.87
	P	0.75(0.84)	0.66	0.10	0.50(0.67)	0.60	0.51	0.50(0.67)	0.72	0.83
	SNPs (n)	46			8			2		
HF	OR	2.17	1.56	32.24	1.35	1.38	1.38	0.52	-	-
	95% CI	1.33-3.55	0.92-2.65	1.28-809.23	1.17-1.56	1.17-1.63	1.00-1.91	0.08-3.14	-	-
	P	2.02×10^{-3} (0.02)*	0.096	0.04	5.24×10^{-5}(9.17×10^{-4})**	1.37×10^{-4}	0.10	0.47(0.67)	-	-
	SNPs (n)	53			8			3		
CD	OR	2.12	2.08	26.38	0.92	0.91	1.06	1.57	1.35	1.07
	95% CI	0.95-4.75	0.67-6.47	0.19-3727.86	0.66-1.27	0.62-1.35	0.52-2.15	0.09-26.67	0.05-37.94	0.00-5324.96
	P	0.07(0.23)	0.21	0.20	0.61(0.77)	0.65	0.87	0.75(0.84)	0.86	0.99

	SNPs (n)	53			8			3		
AA	OR	2.92	1.77	0.04	1.59	1.70	1.69	0.16	0.06	183.83
	95% CI	0.84-10.16	0.32-9.90	0-76.96	0.79-3.19	0.95-3.04	0.34-8.35	0.002-11.58	0-12.58	0.001-6730 2153.04
	P	0.09(0.23)	0.52	0.40	0.19(0.40)	0.07	0.54	0.39(0.62)	0.30	0.57
stroke	SNPs (n)	53			8			3		
	OR	1.55	1.21	0.82	1.10	1.09	1.03	0.15	0.15	0.22
	95% CI	0.87-2.76	0.64-2.26	0.02-37.45	0.88-1.37	0.91-1.32	0.63-1.69	0.04-0.62	0.03-0.80	0.003-15.22
	P	0.14(0.30)	0.56	0.92	0.43(0.65)	0.36	0.90	8.66×10^{-3} (0.04)*	0.03	0.61
tachycardia	SNPs (n)	25			-			2		
	OR	1.00	1.00	0.99	-	-	-	1.00	-	-
	95% CI	1.00-1.01	0.99-1.01	0.93-1.05	-	-	-	0.98-1.02	-	-

	P	0.38(0.62)	0.80	0.72	-	-	-	0.89(0.91)	-	-
bradycardia	SNPs(n)	37			3			2		
	OR	0.99	1.00	0.99	1.00	1.00	1.00	0.99	-	-
	95% CI	0.99-1.00	0.99-1.01	0.94-1.04	1.00-1.00	1.00-1.00	0.99-1.00	0.96-1.01	-	-
	P	0.08(0.23)	0.52	0.69	0.54 (0.71)	0.51	0.71	0.32(0.59)	-	-
DCM	SNPs(n)	52			8			3		
	OR	5.48	13.59	>100	1.70	1.67	2.48	2.30	0.12	0.005
	95% CI	0.82-36.56	1.49-123.86	115.19-943382652804.39	0.93-3.09	0.83-3.37	0.72-8.58	0.01-1021.89	0.00-132.69	0.00–639458561.05
	P	0.08(0.23)	0.02	0.008	0.09(0.23)	0.15	0.20	0.79(0.84)	0.56	0.75

Values are OR (95% CI); P values. FDR-adjusted q values are shown only for IVW estimates. IVW, inverse-variance weighted; WM, weighted median; OR, odds ratio; CI, confidence interval; FDR, Benjamini–Hochberg false discovery rate. FDR correction was applied across all 35 estimable IVW exposure–outcome tests (Cigarettes per day–tachycardia could not be estimated because only 2 instruments were available after harmonisation). Bonferroni threshold = 0.05/35 = 1.43×10^{-3}. ** P < Bonferroni threshold; * q < 0.05. AF, atrial fibrillation; CAD, coronary artery disease; DVT/PE,

deep venous thrombosis and pulmonary embolism; PAD, peripheral artery disease; HF, heart failure; CD, death due to cardiac causes; AA, aortic aneurysm; VA, ventricular arrhythmia; DCM, dilated cardiomyopathy.

Sensitivity Analyses

MR-PRESSO detected horizontal pleiotropy in five exposure–outcome pairs (Supplementary Table S2). Outlier correction left the "ever smoked"–atrial fibrillation and "cigarettes per day"–PAD estimates essentially unchanged (OR 1.27 → 1.26 and OR 1.74 → 1.72, respectively), whereas several other estimates shifted. For the "ever smoked"–CAD association, correction of two outlier SNPs attenuated the OR from 2.84 to 2.00 but remained significant (P = 0.005). For "ever smoked"–DCM, pleiotropy was detected and outlier correction shifted the estimate from OR = 5.48 (P = 0.085) to OR = 7.78 (P = 0.029); this association should therefore be interpreted with caution. For "exposure to tobacco smoke outside home"–atrial fibrillation, correction of one outlier reversed the direction of the estimate (OR 1.42 → 0.95), consistent with the confounded nature of this UVMR association noted above. Steiger filtering confirmed correct causal directionality for all computable exposure – outcome pairs (Supplementary Table S3). Leave-one-out analyses demonstrated that the core associations for "ever smoked"–CAD, HF, and PAD were stable, while the DVT/PE association was mostly stable; several null associations were unstable (Supplementary Table S4).

MVMR Analysis

To estimate the direct effects of smoking behaviours on cardiovascular outcomes independent of cardiometabolic mediators, MVMR was restricted to the five exposure–outcome pairs that survived FDR correction in the univariable MR analysis: Ever smoked–CAD, Ever smoked–HF, Ever smoked–PAD, Cigarettes per day–PAD, and

Cigarettes per day–HF. All models were fitted using IVW with robust standard errors.

Ever smoked: The direct effect of "ever smoked" on CAD persisted after adjustment for each cardiometabolic mediator individually and in combination: T2DM (OR = 2.62, 95% CI 1.50–4.58, P = 1.4×10^{-3}), hypertension (OR = 3.32, 95% CI 2.03–5.41, P = 1.3×10^{-5}), BMI (OR = 2.66, 95% CI 1.40–5.07, P = 4.7×10^{-3}), HDL-C (OR = 2.86, 95% CI 1.74–4.71, P = 1.3×10^{-4}), LDL-C (OR = 2.50, 95% CI 1.50–4.17, P = 9.0×10^{-4}), TG (OR = 3.03, 95% CI 1.80–5.12, P = 1.2×10^{-4}), Lp(a) (OR = 2.58, 95% CI 1.11–6.02, P = 0.035), the metabolic model (T2DM + BMI; OR = 2.54, 95% CI 1.32–4.90, P = 7.7×10^{-3}), the lipid model (LDL-C + TG + HDL-C; OR = 2.79, 95% CI 1.69–4.61, P = 2.0×10^{-4}), and the all-available model (OR = 2.96, 95% CI 1.68–5.21, P = 5.3×10^{-4}). Directionally consistent and statistically significant direct effects were also observed for HF and PAD across all single-mediator and combined models (HF: OR range 1.55–2.41 across all models; PAD: OR range 2.42–3.66), with the exception of the Lp(a)-adjusted model for PAD, in which the estimate was directionally consistent but did not reach significance (OR = 2.10, 95% CI 0.73–6.02, P = 0.18). Conditional F-statistics exceeded 10 in all single-mediator and metabolic/lipid models; the all-available model (T2DM + BMI + hypertension + LDL-C + TG) had a minimum conditional F-statistic of 10.0 and was therefore reported as exploratory.

Cigarettes per day: For the UK Biobank-free "cigarettes per day" instrument set, direct effects on PAD were robust after adjustment for HDL-C (OR = 7.21, 95% CI 4.05–12.84, P = 6.8×10^{-3}) and LDL-C (OR = 8.35, 95% CI 4.79–14.56, P = 4.9×10^{-3}). Direct effects on HF were also significant after adjustment for HDL-C (OR = 1.38, 95% CI 1.16–1.65, P = 0.038) and LDL-C (OR = 1.42, 95% CI 1.15–1.74, P = 0.045). Models adjusting for T2DM, hypertension, or TG, and the combined lipid model, were directionally consistent for PAD (OR range 5.68–7.39) but had conditional F-statistics below 10 and were flagged as exploratory; the corresponding HF models were null (OR range 1.11–1.44, all P > 0.05) and also exploratory. These findings should be interpreted

with caution given the limited number of instruments (n = 5) available for this exposure after confounder screening.

Overall, the MVMR analyses confirmed that the direct effects of "ever smoked" on CAD, HF, and PAD persisted after adjustment for individual and combined cardiometabolic risk factors, and that the direct effects of "cigarettes per day" on PAD and HF persisted after adjustment for HDL-C and LDL-C. No MVMR analyses were performed for DVT/PE, DCM, or the "exposure to tobacco smoke outside home"–CAD association because these UVMR associations did not survive FDR correction. Full MVMR results are presented in Table 3, Supplementary Table S5 and Figure 2.

Table 3 Multivariable MR direct effects after adjustment for cardiometabolic confounders.

Exposures	Adjusted for	Outcomes	SNP (n)	OR (95% CI)	P	Cond. F	Robustness
ever smoked	T2DM	CAD	56	2.62 (1.50-4.58)	1.4×10^{-3}	29.3	Robust
	Hypertension		56	3.32 (2.03-5.41)	1.3×10^{-5}	30.8	Robust
	BMI		47	2.66 (1.40-5.07)	4.7×10^{-3}	31.4	Robust
	HDL-C		56	2.86 (1.74-4.71)	1.3×10^{-4}	35.6	Robust
	LDL-C		56	2.50 (1.50-4.17)	9.0×10^{-4}	34.6	Robust
	TG		56	3.03 (1.80-5.12)	1.2×10^{-4}	30.8	Robust
	Lp(a)		34	2.58 (1.11-6.02)	0.04	31.0	Robust
	T2DM+BMI		47	2.54 (1.32-4.90)	7.7×10^{-3}	22.2	Robust
	LDL-C+TG+HDL-C		56	2.79 (1.69-4.61)	2.0×10^{-4}	26.0	Robust
	T2DM+BMI+Hypertension+LDL-C+TG		47	2.96 (1.68-5.21)	5.3×10^{-4}	10.0	Exploratory
	T2DM	HF	56	2.01 (1.29-	$3.1\times10^{-}$	29.3	Robust

			3.13)	^{3}		
Hypertensio n		56	2.41 (1.60-3.62)	9.4×10^{-5}	30.8	Robust
BMI		47	1.89 (1.15-3.09)	0.02	31.4	Robust
HDL-C		56	2.18 (1.42-3.33)	7.0×10^{-4}	35.6	Robust
LDL-C		56	1.97 (1.31-2.95)	1.8×10^{-3}	34.6	Robust
TG		56	2.12 (1.35-3.31)	1.8×10^{-3}	30.8	Robust
Lp(a)		34	2.41 (1.33-4.37)	6.6×10^{-3}	31.0	Robust
T2DM+BM I		47	1.75 (1.08-2.84)	0.03	22.2	Robust
LDL-C+TG+HD L-C		56	1.95 (1.29-2.94)	2.6×10^{-3}	26.0	Robust
T2DM+BM I+Hyperten sion+LDL-C+TG		47	1.55 (1.02-2.35)	0.045	10.0	Explorat ory
T2DM	PAD	56	2.50 (1.20-5.21)	0.02	29.3	Robust
Hypertensio n		56	3.66 (1.89-7.07)	3.0×10^{-4}	30.8	Robust
BMI		47	2.64 (1.24-5.62)	0.02	31.4	Robust
HDL-C		56	2.87 (1.45-5.67)	3.7×10^{-3}	35.6	Robust
LDL-C		56	2.42 (1.27-4.64)	0.01	34.6	Robust
TG		56	3.06 (1.57-5.96)	1.7×10^{-3}	30.8	Robust
Lp(a)		34	2.10 (0.73-6.02)	0.18	31.0	Robust
T2DM+BM I		47	2.63 (1.22-5.70)	0.02	22.2	Robust
LDL-C+TG+HD L-C		56	2.81 (1.53-5.17)	1.7×10^{-3}	26.0	Robust
T2DM+BM I+Hyperten sion+LDL-		47	3.51 (1.76-7.00)	9.3×10^{-4}	10.0	Explorat ory

	C+TG						
Cigarettes per day	HDL-C	PAD	5	7.21 (4.05-12.84)	6.8×10^{-3}	13.4	Robust
	LDL-C		5	8.35 (4.79-14.56)	4.9×10^{-3}	61.1	Robust
	T2DM		5	6.51 (3.15-13.45)	0.02	3.82	Exploratory
	Hypertension		5	5.68 (3.31-9.76)	8.1×10^{-3}	8.04	Exploratory
	TG		5	5.85 (3.07-11.18)	0.01	3.29	Exploratory
	LDL-C+TG+HDL-C		5	7.39 (6.18-8.82)	0.03	2.62	Exploratory
	HDL-C	HF	5	1.38 (1.16-1.65)	0.04	13.4	Robust
	LDL-C		5	1.42 (1.15-1.74)	0.045	61.1	Robust
	T2DM		5	1.11 (0.88-1.39)	0.44	3.82	Exploratory
	Hypertension		5	1.17 (0.84-1.62)	0.41	8.04	Exploratory
	TG		5	1.26 (0.87-1.83)	0.31	3.29	Exploratory
	LDL-C+TG+HDL-C		5	1.44 (1.18-1.74)	0.17	2.62	Exploratory

MVMR was restricted to exposure–outcome pairs that survived FDR correction in the univariable MR analysis. OR, odds ratio; CI, confidence interval; IVW, inverse-variance weighted; T2DM, type 2 diabetes mellitus; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TG, triglycerides; Lp(a), lipoprotein A; BMI, body mass index; CAD, coronary artery disease; HF, heart failure; PAD, peripheral artery disease. ORs represent the direct effect of the exposure on the outcome after adjustment for the indicated mediator(s). All estimates were obtained using IVW with robust standard errors. "Cond. F" denotes the minimum conditional F-statistic across all instruments in the model; models with any conditional F < 10 are flagged as "Exploratory".

Fig. 2. Multivariable Mendelian randomization direct effects of smoking behaviours on cardiovascular outcomes after adjustment for cardiometabolic confounders. Forest plots show IVW odds ratios (ORs) with 95% confidence intervals for (A) Ever smoked → CAD, (B) Ever smoked → HF, (C) Ever smoked → PAD, (D) Cigarettes per day → PAD, and (E) Cigarettes per day → HF. Each row represents a separate MVMR model adjusting for the indicated mediator(s): "Metabolic" = T2DM + BMI; "Lipids" = LDL-C + TG + HDL-C; "All" = T2DM + BMI + hypertension + LDL-C + TG. Blue, models with all conditional F-statistics > 10; orange, models with at least one conditional F < 10 (exploratory). The dashed vertical line indicates the null (OR = 1). MVMR, multivariable Mendelian randomization; CAD, coronary artery disease; HF, heart failure; PAD, peripheral artery disease; T2DM, type 2 diabetes mellitus; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TG, triglycerides; Lp(a), lipoprotein A; BMI, body mass index.

eQTL-SMR Analyses

After excluding HLA region genes, eQTL-SMR identified 4,278 unique genes across the five CVD subtypes (CAD: 1,420; DCM: 709; DVT/PE: 917; PAD: 866; HF: 1,097). A total of 640 genes were shared by at least two outcomes: 593 by exactly two, 43 by three, and 4 by four; no gene reached significance across all five subtypes (Figure 3A). KEGG enrichment of the shared genes identified eight pathways with adjusted $p < 0.05$ (Figure 3B): insulin resistance (adjusted $p = 0.014$), cellular senescence (adjusted $p = 0.039$), Yersinia infection (adjusted $p = 0.039$), peroxisome (adjusted $p = 0.046$), Salmonella infection (adjusted $p = 0.046$), Toll-like receptor signaling (adjusted $p = 0.046$), HIF-1 signaling (adjusted $p = 0.046$), and toxoplasmosis (adjusted $p = 0.046$). These pathways converged on metabolic dysregulation, inflammatory/immune activation, and cellular stress. Insulin signaling (adjusted $p = 0.055$) and chemokine signaling (adjusted $p = 0.058$) were borderline and were not considered significant under BH-FDR. GO BP enrichment did not identify any pathway with adjusted $p < 0.05$ (Figure 3C). At the nominal level, the top terms included negative regulation of ER-unfolded protein response ($P = 1.4 \times 10^{-4}$), regulation of protein polymerization ($P =$

1.4 × 10^{-4}), and regulation of leukocyte migration (P = 2.9 × 10^{-4}); these were considered exploratory and are reported for hypothesis generation only.

Fig. 3. Functional characterization of SMR-significant genes shared across CVD subtypes. (A) UpSet plot of SMR-significant genes across five CVD subtypes after excluding HLA region genes. A total of 640 genes were shared by at least two outcomes (593 by two, 43 by three, and 4 by four); no gene was shared by all five subtypes. (B) KEGG pathways enriched in shared SMR genes. Eight pathways survived Benjamini – Hochberg correction (adjusted $p < 0.05$), converging on metabolic dysregulation, immune/inflammatory activation, and cellular stress. The dashed line indicates adjusted $p = 0.05$. (C) GO biological processes nominally enriched in shared SMR genes. No term survived Benjamini–Hochberg correction (all adjusted $p > 0.15$); nominal P values are shown for hypothesis generation.

Colocalization of smoking-associated DNA methylation with cis-eQTLs

To link smoking-induced DNA methylation changes to the shared transcriptional programs, we tested 296 CpG-gene pairs in which a smoking-associated CpG was located within ±100 kb of one of the 640 shared eQTL-SMR genes (Supplementary Table S10). Using GoDMC cis-mQTL and eQTLGen cis-eQTL data, we performed Bayesian colocalization. Twenty-five pairs showed PP.H4 > 0.75. To distinguish multi-SNP from single-SNP driven colocalization, we removed the lead mQTL and eQTL SNPs and re-ran coloc. Eleven pairs retained PP.H4 > 0.75, involving seven genes: SPTLC2 (cg14544289), AMDHD1 (cg00481159), FADS2 (cg01400685), RPS6KL1 (cg06998765 and cg16124934), PROB1 (cg08719608 and cg08220120), POLR2A (cg10001987 and cg27579313), and HIC1 (cg20664636 and cg01160692) (Figure 4; Supplementary Table S11). An extension analysis of the nearest three CpGs per gene identified three additional colocalized CpGs (PROB1 cg08220120, POLR2A cg27579313, HIC1 cg01160692), all supported by multiple SNPs. Two pairs involving CAMK1D (cg03575602 and cg14537549) showed complete colocalization driven by a

single cis variant (rs11257655), which was the lead SNP for both mQTL and eQTL. One pair involving SPTLC2 (cg20385229) showed suggestive colocalization (PP.H4 = 0.825) but with limited SNP coverage (n = 76) (Supplementary Figure S1). Functional annotation of the candidate genes identified 29 enriched terms (3 KEGG pathways and 26 GO biological processes), including lipid metabolism (FADS2, SPTLC2), oxidative stress and inflammatory signaling (CAMK1D, FADS2), and amino acid metabolism (AMDHD1), consistent with established smoking-related pathways (Supplementary Table S12). Notably, AHRR, the most established smoking methylation marker, did not colocalize with its cis-eQTL (all PP.H4 < 0.06), indicating that not all smoking-associated CpGs mediate their effects through shared causal variants with gene expression (Supplementary Table S10).

Fig. 4. Candidate axes linking smoking-associated DNA methylation to shared eQTL-SMR genes and cardiovascular disease subtypes. Sankey diagram showing one representative CpG per gene (highest PP.H4), the colocalized gene, and the CVD subtypes for which the gene reached nominal significance (p_SMR < 0.05) in eQTL-SMR analysis. Only CpG-gene pairs with multi-SNP or single-SNP colocalization are shown. Flow width is proportional to the number of CpG-gene-subtype links. DCM was excluded because it did not survive FDR correction in univariable MR. CpG, 5′-cytosine-phosphate-guanine-3′; eQTL-SMR, summary-data-based Mendelian randomization using expression quantitative trait loci; CVD, cardiovascular disease; CAD, coronary artery disease; HF, heart failure; PAD, peripheral artery disease; DVT/PE, deep venous thrombosis/pulmonary embolism; PP.H4, posterior probability of a shared causal variant.

Decoupling of epigenetic imprints and gene expression

To link smoking-induced DNA methylation changes to the shared transcriptional programs, we tested 296 CpG-gene pairs. Twenty-five pairs showed PP.H4 > 0.75, and eleven pairs retained this significance after removing lead SNPs, involving seven genes. However, this colocalization was not universal. The most established smoking

methylation marker, AHRR, showed complete absence of mQTL-eQTL colocalization (maximum PP.H4 = 0.059; PP.H3 = 0.94–1.00 across 47 CpG-gene pairs). Further investigation at the AHRR locus revealed two independent cis-mQTL signals. The strongest mQTL lead SNP (rs11742897, $P = 3.8\times10^{-92}$) showed no association with AHRR expression (eQTL P = 0.74) (Supplementary Figure S2 and Supplementary Table S13), while the second SNP (rs2466287) showed opposite effects on methylation and expression. This allelic heterogeneity indicates that the classical smoking methylation biomarkers reflect exposure rather than serving as causal mediators of gene expression.

pQTL MR Analyses

pQTL MR identified 18 protein–CVD associations surviving FDR correction (FDR < 0.05), distributed across four subtypes: 6 for CAD, 6 for DVT/PE, 3 for HF, and 3 for PAD (Figure 5 and Supplementary Table S6). Notably, no proteins were shared across subtypes. The strongest signal was ABO with DVT/PE ($\beta = 0.0036$, SE = 0.00021, $P = 2.1\times10^{-68}$, $FDR = 3.8\times10^{-65}$). For CAD, top signals were Cardiotrophin-1 ($\beta = 0.134$, $P = 9.3\times10^{-23}$) and VPS29 ($\beta = -0.079$, $P = 1.5\times10^{-16}$). For HF, Granulins ($\beta = 0.063$, $P = 2.7\times10^{-9}$) and IL-3RA ($\beta = -0.061$, $P = 4.2\times10^{-8}$) were significant. For PAD, Haptoglobin ($\beta = -0.096$, $P = 3.6\times10^{-8}$) and Ferritin ($\beta = -0.138$, $P = 1.8\times10^{-6}$) showed protective associations. All 18 proteins showed consistent effect directions across IVW, weighted median, and MR-Egger.

Fig. 5. Proteome-wide pQTL MR identifies 18 causal proteins for CVD subtypes. Forest plot shows IVW odds ratios (ORs) with 95% confidence intervals per 1-unit increase in plasma protein level (RFU, relative fluorescence units; INTERVAL SomaScan platform). Colors denote CVD subtype: CAD (blue), DVT/PE (red), HF (orange), PAD (purple). The dashed vertical line indicates the null (OR = 1). All 18 proteins survived Benjamini–Hochberg FDR correction (FDR < 0.05) and showed directionally consistent estimates across IVW, weighted median, and MR-Egger. Because SomaScan RFU values have a narrow dynamic range, several DVT/PE effect

sizes are numerically small on the OR scale despite extreme statistical significance (e.g., ABO: $\beta = 0.0036$, $P = 2.1 \times 10^{-68}$, FDR = 3.8×10^{-65}).

Systemic exclusion of candidate proteins across platforms

We next investigated the eQTL-pQTL continuity for the seven genes with robust mQTL-eQTL colocalization. We found a profound decoupling. While POLR2A was covered by the UKB-PPP Olink panel, its cis-pQTL signal was also sub-genome-wide (minimum $P = 1.95\times10^{-4}$), precluding formal MR analysis. In the INTERVAL plasma dataset, the only candidate gene measured was SPTLC2, yet its eQTL-pQTL colocalization was dominated by PP.H1 (0.65), indicating an eQTL-only signal with no shared causal variant for plasma protein levels (minimum pQTL $P = 3.88\times10^{-5}$) (Supplementary Figure S3 and Supplementary Table S14).

To determine whether this was a plasma-specific phenomenon, we interrogated the SomaScan 7k CSF dataset (NIAGADS NG00130). SPTLC2 was measured (SeqId 7737-76), but its cis-pQTL signal remained sub-genome-wide (minimum $P = 4.96\times10^{-4}$; max -log10(P) = 3.30 across 6,682 cis-SNPs). Furthermore, a cross-platform coverage analysis against the SomaLogic v4.1 official menu (7,267 analytes) revealed that 6 out of the 7 candidate genes were completely absent from the platform's design. In contrast, all four positive control proteins (ABO, CTF1, GRN, IL3RA) were represented with multiple redundant probes (6–37 per protein) and robust pQTL signals (Figure 6 and Supplementary Table S15).

Fig. 6. Cross-platform coverage and detection limits of candidate proteins. Heatmap illustrating the presence and pQTL signal strength of the seven candidate genes (top) and four positive control proteins (bottom) across the SomaLogic 7k official menu, INTERVAL plasma SomaScan, UKB-PPP plasma Olink, and NIAGADS CSF SomaScan datasets. While positive controls (ABO, CTF1, GRN, IL3RA) show robust detection and strong pQTL signals across platforms, six of the seven candidate genes are completely absent from the SomaLogic menu. The sole covered candidate gene, SPTLC2, exhibits only sub-genome-wide weak pQTL signals in both plasma

(INTERVAL) and cerebrospinal fluid (CSF).

Smoking-to-protein MR Analyses

We next examined whether the 18 FDR-significant CVD-causal proteins act as downstream effectors of smoking exposure by performing two-sample MR using the four smoking behaviours as exposures and these proteins as outcomes (Supplementary Table S7). After FDR correction within each smoking behaviour, no protein reached statistical significance (all FDR > 0.59). Only one nominal association was observed: exposure to tobacco smoke outside home with EVA1C ($\beta = -6.11$, SE = 2.40, P = 0.011, FDR = 0.593), which was based on only three instruments and did not survive multiple-testing correction. Notably, the strongest pQTL signal, ABO, showed a directionally consistent but non-significant association with cigarettes per day ($\beta = 0.236$, SE = 0.132, P = 0.075, FDR = 0.801). These findings confirm a fundamental decoupling: the circulating protein effectors driving CVD subtype risk are independent of direct smoking-induced alterations, suggesting that smoking increases cardiovascular risk through pathways parallel to, rather than linearly upstream of, the circulating proteome.

Colocalization

At the CHRNA3/CHRNA5 locus (chr15:78.8 Mb), Bayesian colocalization revealed an independent CAD-specific signal for "ever smoked"–CAD (PP.H2 = 0.859, PP.H4 = 0.002), indicating that the smoking–CAD association at this locus is not driven by a shared causal variant. In contrast, for PAD, the posterior probability of a shared causal variant was 0.710, suggesting a likely shared genetic driver. Colocalization for HF was inconclusive (PP.H4 = 0.065), with insufficient signal in the region (Supplementary Table S8).

Pathway-level convergence despite gene-level decoupling

Functional convergence was evident at the pathway level despite limited gene-level overlap (Supplementary Table S9). SMR identified three functional axes common across CVD subtypes: metabolic dysregulation, inflammatory activation, and cellular stress. Integration with pQTL enrichment added a fourth axis, complement/coagulation. Pathway enrichment of the 18 proteins converged on the same axes: complement

cascade (Reactome, P = 0.0023), IL-6 family signalling (Reactome, P = 0.030), neutrophil degranulation (Reactome, P = 0.020), cellular oxidant detoxification (GO, P = 0.003), ferroptosis (KEGG, P = 0.044).

While specific linear molecular bridges are lacking, these layers represent a systemic convergence. For instance, ABO maps to blood-group biosynthesis, SELENOS maps to ER stress/autophagy pathways from SMR, and CTF1 maps to IL-6 family signalling. Together, these multi-omic layers define a refined model: smoking-induced DNA methylation converges on shared upstream transcriptional programs, while independent, subtype-specific protein effectors operate downstream. The two layers are biologically coupled at the pathway level but linearly decoupled at the gene-specific level (Figure 7).

Fig. 7. Multi-omics decoupling model linking smoking exposure to CVD subtypes. Shared upstream transcriptional programs (from eQTL-SMR) converge on three functional axes (metabolic dysregulation, inflammatory activation, and cellular stress). However, a systemic decoupling exists between the transcriptome and the circulating proteome (e.g., failure of eQTL-pQTL colocalization for candidate genes). Downstream subtype-specific protein effectors are identified independently by pQTL MR. Notably, these effectors are not direct downstream targets of smoking (Smoking-to-protein MR negative). Despite gene-level decoupling, macroscopic pathway-level convergence remains evident across four functional axes, defining a parallel architecture rather than a linear cascade.

Discussion

This integrative multi-omics MR study provides a comprehensive dissection of the causal effects of smoking behaviours on a broad spectrum of CVDs. Our findings demonstrate that smoking behaviours exert heterogeneous effects on CVDs through a shared upstream transcriptional program with subtype-specific downstream protein effectors. However, rather than a linear molecular cascade, our data reveal a multi-layer decoupled architecture. Five key observations emerge: (1) Direct effects of smoking on

CAD, HF, PAD, and DVT/PE persist after adjustment for all traditional cardiometabolic confounders, indicating that metabolic pathways cannot fully explain the smoking-CVD association; (2) eQTL-SMR identifies shared transcriptional programs across CVD subtypes, converging on metabolic dysregulation, inflammatory activation, and cellular stress; (3) the classical smoking methylation marker AHRR shows independent genetic architectures for its mQTL and eQTL, indicating that epigenetic imprints do not linearly drive gene expression (Decoupling Layer 1); (4) a systemic decoupling exists between the transcriptome and the circulating proteome, driven by the exclusion of intracellular proteins from current detection platforms (Decoupling Layer 2); and (5) colocalization at the CHRNA3/CHRNA5 locus supports a shared causal variant for PAD but an independent signal for CAD, reinforcing the notion that smoking affects different vascular beds through distinct genetic architectures. These findings extend prior observational and MR evidence by moving beyond simple causal inference to map the molecular pathways linking smoking to cardiovascular morbidity, with direct implications for phenotype-specific prevention.

Our UVMR analyses demonstrated that genetic predisposition to 'ever smoked' increases the risk of CAD, DCM, and DVT/PE, whereas 'cigarettes per day' specifically elevates the risk of PAD and HF. These findings reinforce the concept that smoking is a heterogeneous exposure: the mere initiation of smoking ('ever smoked') is sufficient to confer risk for CAD, DCM, and DVT/PE, whereas PAD and HF appear to be more sensitive to smoking intensity—implying that dose-response relationships differ by vascular bed and myocardial phenotype. This pattern extends prior meta-analytic evidence [11], which demonstrated that even light smoking carries about half the excess risk of CAD and stroke seen in heavy smoking.

No causal associations were observed for AF, tachycardia, or bradycardia across any smoking phenotype. The absence of AF association may reflect that our genetic instruments for 'ever smoked' include former smokers, potentially diluting an effect that attenuates after cessation [12]. Furthermore, while nicotine acutely altered cardiac electrophysiology, lifelong genetic predisposition may preferentially capture chronic

structural and metabolic consequences rather than transient arrhythmogenic effects. Similarly, a dose-dependent relationship between smoking and venous thromboembolism has been reported [13]. Regarding DVT/PE, we detected a significant but modest effect of ‘ever smoked’, yet no effect of ‘cigarettes smoked per day’, indicating that dose-dependent relationships reported observationally may not be fully captured by existing genetic instruments for smoking intensity. The null finding for AA is notable given the well-established association between smoking and abdominal aortic aneurysm (AAA), with meta-analyses reporting a nearly 5-fold increased risk in current smokers [14]. This discrepancy may reflect site-specific heterogeneity: smoking is strongly associated with AAA but not with thoracic aortic aneurysm (TAA) [15]. Our composite AA phenotype may have diluted the specific AAA signal. Additionally, MR captures lifetime genetic liability to smoking, whereas observational studies measure cumulative active exposure, which may better reflect long-term vascular damage. Subtype-specific AA GWAS are needed to clarify this finding. Furthermore, for 'exposure to tobacco smoke outside home', the suggestive associations with stroke and VA in UVMR were highly unstable. The stroke association was completely reversed after adjusting for BMI, indicating confounding rather than a genuine causal effect, while the VA estimate was imprecise and not supported by sensitivity analyses. Due to this instability and limited instrument availability, neither outcome was carried forward to downstream analyses, and we caution against overinterpreting these exploratory findings.

MVMR disentangled direct and mediated pathways. Our MVMR analyses robustly demonstrated that the adverse effects of smoking on CAD, HF, and PAD are largely direct and independent of traditional cardiometabolic risk factors. Even after comprehensively adjusting for T2DM, hypertension, BMI, and lipid fractions (LDL-C, HDL-C, TG), the direct effects of 'ever smoked' remained highly robust for CAD, HF, and PAD. Specifically, for smoking intensity ('cigarettes per day'), the effects on PAD and HF persisted with surprisingly high magnitudes after adjusting for HDL-C and LDL-C. This indicates that while smoking induces dyslipidemia and dysglycemia, the

atherogenic and cardiotoxic properties of tobacco extend far beyond traditional metabolic disturbances. Consistent with this, a recent multivariable MR separating nicotine from non-nicotine constituents found no clear independent effect of the nicotine metabolite ratio on CHD, attributing the residual risk to non-nicotine combustion products [16]. Thus, the residual direct effects observed here likely reflect direct endothelial injury, systemic inflammation, and oxidative stress [17]. For PAD specifically, although a full MVMR mediation decomposition was constrained by limited genetic instruments (conditional $F < 10$), the direct effect remained robust. While a recent multi-omics study suggested that cotinine-related IL6/STAT3 inflammatory pathways may partially mediate the smoking-PAD association [18], our data show that this inflammatory component is far from sufficient to explain the full effect, given the extreme magnitude of the lipid-adjusted direct effect ($OR > 5$). Notably, this multi-layer decoupling is also evident in the epigenetic layer. Two independent MR studies failed to retain AHRR as a causal CpG for CVDs [19, 20]. Our data further provide the genetic basis for this consistent exclusion: the strongest AHRR mQTL variant had no effect on expression (eQTL $P = 0.74$), ruling out its role as a cis-mediator.

These findings have direct clinical relevance: for PAD and HF, smoking abstinence is irreplaceable; for CAD, because the direct atherogenic effect is strong and independent of lipids and glucose, integrated strategies combining cessation with aggressive management of both dysglycemia and dyslipidemia are likely to yield the greatest benefit. Furthermore, the well-documented reductions in cardiovascular hospitalizations following indoor smoking bans emphasize the public health relevance of limiting all forms of tobacco exposure [21].

The eQTL-SMR analysis revealed that the gene sets identified across all five CVD outcomes were highly overlapping, indicating that the transcriptional regulatory architecture underlying CVD susceptibility is largely common across different clinical manifestations, consistent with shared pathophysiological mechanisms such as inflammation and lipid metabolism. However, integrating smoking-associated DNA methylation with eQTL data revealed a striking decoupling. While 11 CpG-gene pairs

exhibited robust colocalization, the most established smoking methylation biomarker, AHRR (cg05575921), showed a complete absence of mQTL-eQTL colocalization (PP.H3 = 0.94–1.00 across 47 CpG-gene pairs), and its strongest mQTL lead SNP (rs11742897) had no effect on expression (eQTL P = 0.74).

Our observation that AHRR mQTL and eQTL architectures are independent is concordant with recent MR studies. For instance, a colocalization study of smoking and nine CVDs identified five key mediating CpGs but did not retain AHRR as a causal CpG [19], and a UK Biobank–FinnGen–BBJ MR of smoking-related CpGs and CHD/MI similarly failed to retain AHRR among the robust colocalizing sites [20]. Consistent with AHRR acting as a biomarker rather than a mediator, a prospective cohort of 11,332 individuals found that AHRR hypomethylation predicted PAD (SHR = 2.82, 95% CI 1.91–4.15) and AA (SHR = 2.88, 95% CI 1.42–5.88) risk. This predictive capacity reflects exposure imprinting, not a demonstrated causal pathway from methylation to transcription to disease [22].

Our data provide the genetic architecture basis for this consistent exclusion. The AHRR locus exhibits a highly disproportionate trans-regulatory architecture (3 cis-eQTMs vs. 655 trans-eQTMs), meaning a negative cis-colocalization result is mechanistically expected if the true regulatory relationship is trans-mediated. Therefore, our findings clarify that AHRR methylation serves as an exceptionally robust exposure imprint—reflecting smoking exposure with high accuracy—rather than functioning as a causal mediator of downstream transcriptional programs.

The more profound decoupling emerged between the transcriptome and the circulating proteome. Six of the seven candidate genes with robust mQTL-eQTL colocalization were completely absent from the SomaLogic 7k official menu. For SPTLC2, the sole covered gene, eQTL-pQTL colocalization was negative in plasma (PP.H1 = 0.65), and its cis-pQTL remained sub-genome-wide in cerebrospinal fluid (minimum P = 4.96×10^{-4}). Furthermore, while POLR2A was covered by the UKB-PPP Olink panel, its cis-pQTL signal was sub-genome-wide (minimum P = 1.95×10^{-4}),

precluding MR analysis. This indicates that the difficulty is not simply plasma dilution but an intrinsic limitation in capturing membrane-protein regulation in accessible biofluids. This is not a platform-specific artefact but a predictable manifestation of platform-level biological bias. The UKB-PPP consortium quantified that nuclear proteins are significantly depleted among cis-pQTL-bearing proteins (OR = 0.55, P = 2.2×10^{-12}) [23]. More directly, the SomaScan 7k expansion in 13,445 Europeans showed that secreted proteins were more than four times more likely to harbour a cis-pQTL than intracellular proteins (OR = 4.25), and that the reduced cis-pQTL discovery among newly added proteins was largely explained by their biological characteristics—intracellular localization and low circulating abundance—rather than reduced assay sensitivity [24]. SPTLC2 (an ER membrane enzyme) and POLR2A (a nuclear RNA polymerase subunit) fall squarely into this biologically inaccessible range. One apparent tension deserves attention. A joint eQTL–pQTL analysis reported that genes with strong eQTL–pQTL overlap predominantly encode intracellular proteins, whereas weak overlap tends to occur for glycosylated secreted proteins [25]. This might seem to contradict our finding that intracellular proteins are the ones that fail eQTL–pQTL continuity. The reconciliation lies at the level of detectability: when intracellular proteins are reliably measured, their genetic architecture can indeed align more closely with the transcriptome; but most intracellular proteins are not reliably detected in plasma in the first place. The decoupling we observe therefore operates at the prior step of detection accessibility, not at the step of regulation architecture.

Our pQTL MR (using the INTERVAL dataset) identified 18 subtype-specific causal proteins. The strongest signal was ABO with DVT/PE, consistent with its well-established pro-thrombotic role [26]. Mechanistically, ABO determines the glycosylation of von Willebrand factor and factor VIII , a mechanism intrinsic to the haemostatic system and genetically determined by the ABO locus itself. Mendelian randomization analyses formally suggest that high factor VIII levels are causally linked to venous thromboembolism and coronary artery disease risk [27], and the ABO

rs8176719 variant is associated with VTE risk in a large population-based cohort (HR = 1.30, P = 3.9×10^{-10}) [28]. Critically, smoking-to-protein MR showed no significant causal effects of any smoking behaviour on these 18 proteins. This contrasts with a recent proteome-wide two-step MR that identified ASGR1 as a mediator explaining 6.35% (95% CI 2.61–10.09%, P = 8.77×10^{-4}) of smoking's effect on ischemic heart disease, with colocalization support (PP.H4 = 0.91) [29]. However, that study also noted that the ASGR1-IHD residual association was not significant in never-smokers, with BMI implicated as a potential contributing pathway, consistent with possible horizontal pleiotropy [29]. Our failure to recover such mediators is methodologically informative. First, single mediator effect sizes are small (e.g., 6.35%) and may fall below the stringent FDR threshold in proteome-wide screens. Second, our pQTL MR was designed to identify proteins with the strongest causal effects on CVD, not those most strongly regulated by smoking. These divergent findings highlight that disease-specific mediators may carry substantial pleiotropic or constitutional components. Therefore, ABO and the other 17 proteins represent constitutional risk factors that operate alongside, not within, the smoking-CVD causal pathway.

Despite gene-level decoupling, both the upstream eQTL-SMR genes and the downstream 18 pQTL effectors converged on four shared functional axes: metabolic dysregulation, inflammatory activation, cellular stress, and complement/coagulation. This is consistent with the established multi-pathway pathophysiology of smoking-related CVD [30]. The reality of system-level perturbation is independently supported by a proteomic smoking index (pSIN) derived from machine learning on 2,917 plasma proteins in 43,914 UK Biobank participants. The pSIN accurately distinguishes current from never smokers (AUC = 0.95) and replicates in the China Kadoorie Biobank (AUC = 0.91). Crucially, it is significantly associated with the risk of 18 major chronic diseases, including cardiovascular outcomes, independent of self-reported smoking history [31]. For PAD specifically, a multi-omics study integrating NHANES, network toxicology, MR, and single-cell transcriptomics identified IL6 and STAT3 as cotinine-related inflammation mediators with genetically predicted effects on PAD susceptibility

[18]. This aligns with our finding that the PAD effect of 'cigarettes per day' remains robust after lipid adjustment: IL6/STAT3–mediated inflammation may represent one component of the direct effect, but it is unlikely to be the sole pathway given the magnitude of the residual direct effect (OR > 5 in lipid-adjusted models) [18]. These findings are consistent with an omnigenic model [32], in which exposure-associated variants and core disease genes reside at distinct loci. The pSIN captures a broad biological response to smoking, but the specific proteins that carry CVD risk in our pQTL MR (e.g., ABO, CTF1, VPS29) are not necessarily the ones most strongly perturbed by smoking. The shared transcriptional programs represent a system-level response to smoking that is robust and reproducible, while the protein effectors represent constitutional determinants of disease risk that are layered upon, but not generated by, the smoking response. Decoupling therefore occurs at the level of molecular identity—which specific gene or protein carries the signal—rather than at the level of biological system.

Several limitations must be acknowledged. First, our analysis was restricted to European ancestry. Second, eQTLGen and INTERVAL are derived from whole blood and plasma, respectively; tissue-specific regulation in the lung, arterial wall, or liver—the primary targets of smoking—may differ substantially from what is observable in accessible biospecimens. Third, we were unable to obtain full summary statistics for the SomaScan 7k plasma dataset (NIAGADS NG00180) due to file size constraints; however, the P-value-only CSF data and platform coverage analysis sufficiently supported our conclusions. Fourth, the absence of a protein from a platform's menu is informative but does not exclude the possibility of tissue-level genetic regulation; future sequencing-based proteomics in disease-relevant tissues will be necessary. Fifth, while we assessed sample overlap and applied Steiger filtering, weak bidirectional effects cannot be entirely excluded. Finally, the smoking instruments capture lifelong genetic liability rather than cumulative exposure.

Conclusion

In conclusion, distinct smoking behaviours exert heterogeneous causal effects on specific cardiovascular diseases. Rather than a linear cascade, smoking increases cardiovascular risk through a multi-layer decoupled architecture. While smoking leaves broad epigenetic and transcriptomic imprints, these alterations do not linearly translate to the circulating proteome due to the systematic exclusion of intracellular proteins from current detection platforms. Meanwhile, independent subtype-specific protein effectors operate downstream to drive disease risk. These findings highlight the necessity of distinguishing molecular enrichment from shared causal continuity in multi-omic Mendelian randomization studies, and support precision prevention and phenotype-tailored therapeutic targeting in smokers.

Declarations

Ethics approval and consent to participate

As this study involved no primary data collection, separate ethical approval was not required. All datasets used in this study were obtained from publicly available sources with appropriate institutional review board approvals.

Data availability

All GWAS summary statistics used in this study are publicly available through the MRC-IEU OpenGWAS platform (https://gwas.mrcieu.ac.uk/), the GWAS Catalog (https://www.ebi.ac.uk/gwas/), the GSCAN Consortium (https://genome.psych.umn.edu/index.php/GSCAN), the FinnGen Consortium (https://r9.finngen.fi/), DIAMANTE/DIAGRAM (https://diagram-consortium.org/downloads.html), GLGC (https://csg.sph.umich.edu/willer/public/glgc-lipids2021/), ICBP, and GIANT (https://giant-consortium.web.broadinstitute.org/). Blood cis-mQTL summary statistics were obtained from the GoDMC consortium (http://api.godmc.org.uk/). Blood cis-eQTL summary statistics are available from the eQTLGen Consortium (https://www.eqtlgen.org/). Plasma cis-pQTL summary statistics were obtained from the INTERVAL SomaScan dataset (Sun et al. 2018; N = 3,301) via the IEU OpenGWAS platform (https://gwas.mrcieu.ac.uk/). Additional proteomic datasets used for cross-

platform coverage analysis are available from the following sources: UKB-PPP via Synapse (https://www.synapse.org/Synapse:syn51364943); SomaLogic official menus via the SomaScan.db Bioconductor package; and the cerebrospinal fluid (CSF) pQTL data (NG00130) via the NIAGADS Knowledge Portal (https://dss.niagads.org/datasets/ng00130/). LD reference data were from the 1000 Genomes Project European reference panel (https://www.internationalgenome.org/). Analysis scripts are available from the corresponding author upon reasonable request. No new data were generated in this study.

Pre-registration

This study was not pre-registered.

Supplementary Materials

Supplementary Table S1. Screening of instrumental SNPs against five behavioural and social confounders for smoking behaviours.

Supplementary Table S2. MR-PRESSO sensitivity analysis for smoking behaviours and cardiovascular outcomes.

Supplementary Table S3. Steiger filtering directionality analysis for smoking behaviours and cardiovascular outcomes.

Supplementary Table S4. Leave-one-out sensitivity analysis for smoking behaviours and cardiovascular outcomes.

Supplementary Table S5. Multivariable Mendelian randomization direct effects after adjustment for cardiometabolic confounders.

Supplementary Table S6. Proteome-wide pQTL MR associations surviving FDR correction.

Supplementary Table S7. Smoking-to-protein Mendelian randomization for the 18 FDR-significant CVD-causal proteins.

Supplementary Table S8. Bayesian colocalization at the CHRNA3/CHRNA5 locus.

Supplementary Table S9. Integration of eQTL-SMR and pQTL MR findings across functional axes and molecular bridges.

Supplementary Table S10. Full colocalization results for 296 CpG-gene pairs.

Supplementary Table S11. Candidate axes linking smoking-associated DNA methylation to shared eQTL-SMR genes.

Supplementary Table S12. KEGG and GO biological process enrichment of the candidate genes.

Supplementary Table S13. Allelic heterogeneity and Wald ratio analysis at the AHRR locus.

Supplementary Table S14. Bayesian colocalization results for SPTLC2 eQTL and pQTL in plasma.

Supplementary Table S15. Cross-platform coverage matrix for candidate and control proteins.

Supplementary Figure S1. Sankey diagram including all candidate axes (14 CpG-gene pairs), including the suggestive pair (cg20385229-SPTLC2, PP.H4 = 0.825, n SNP = 76).

Supplementary Figure S2. Allelic heterogeneity at the AHRR locus. Forest plot displaying the effect sizes of two independent lead SNPs on AHRR DNA methylation (mQTL, red circles) and gene expression (eQTL, blue triangles). The strongest mQTL SNP (rs11742897) has no significant effect on expression (P = 0.74), while rs2466287 exhibits opposite effect directions. This confirms independent genetic architectures and a lack of shared causal variants.

Supplementary Figure S3. eQTL-pQTL colocalization for SPTLC2 in plasma. Stacked bar plot showing the posterior probabilities for the colocalization hypotheses (H0–H4) in the INTERVAL plasma dataset. The signal is overwhelmingly dominated by H1 (eQTL only, PP.H1 = 65.2%), indicating that the genetic variants driving SPTLC2 expression do not share a causal variant with those influencing its plasma protein levels (PP.H4 = 1.1%).

Competing interests

The authors declare no competing interests.

Primary funding

The study was conducted independently without any involvement or influence from

any of the funding agencies. The study was funded by Tianjin Key Medical Discipline Construction Project (TJYXZDXK-3-017B), Tianjin Health Science and Technology Project (TJWJ2023QN063), The Tianjin Traditional Chinese Medicine and Integrated Traditional Chinese & Western Medicine Research Project, supported by the Tianjin Municipal Health Commission (2025150), Tianjin Science Popularization Project (25KPXCRC00430) and Li • Xin Traditional Chinese Medicine Research Innovation Fund (2025-CCA-TCM-072).

Author contributions

Formal analysis, Methodology, Visualization, Software, Data curation, Investigation, Writing –original draft: Shuo Ying, and Qiang Chen; Validation, Visualization, Writing – original draft: Xiaogang Liu. Conceptualization, Supervision, Writing – review & editing: Peng Zhang, Naikuan Fu and Yingwu Liu.

Acknowledgments

We gratefully acknowledge the investigators and participants of the numerous genome-wide association studies whose summary statistics were used in this study. We thank the MRC-IEU OpenGWAS project, the GSCAN Consortium, the FinnGen Consortium, DIAMANTE/DIAGRAM, the Global Lipids Genetics Consortium (GLGC), the International Consortium for Blood Pressure (ICBP), and the GIANT Consortium for making GWAS summary statistics publicly available. We also thank the eQTLGen Consortium for providing blood cis-eQTL data, the INTERVAL study and the IEU OpenGWAS platform for providing the SomaScan pQTL summary statistics used in the pQTL MR analyses, and the 1000 Genomes Project for the linkage disequilibrium reference panel. This research would not have been possible without the commitment of all study participants and the research teams who generated and shared these valuable genetic and molecular resources.

References

1. Mensah GA, Fuster V, Roth GA. A Heart-Healthy and Stroke-Free World: Using Data to Inform Global Action. J Am Coll Cardiol. 2023;82(25):2343-9. doi: 10.1016/j.jacc.2023.11.003.
2. World Health Organization. Tobacco 2025 [Available from: https://www.who.int/en/news-room/fact-sheets/detail/tobacco.
3. Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, et al. 21st-

century hazards of smoking and benefits of cessation in the United States. N Engl J Med. 2013;368(4):341-50. doi: 10.1056/NEJMsa1211128.
4. Siasos G, Tsigkou V, Kokkou E, Oikonomou E, Vavuranakis M, Vlachopoulos C, et al. Smoking and atherosclerosis: mechanisms of disease and new therapeutic approaches. Curr Med Chem. 2014;21(34):3936-48. doi: 10.2174/092986732134141015161539.
5. Dahdah A, Jaggers RM, Sreejit G, Johnson J, Kanuri B, Murphy AJ, et al. Immunological Insights into Cigarette Smoking-Induced Cardiovascular Disease Risk. Cells. 2022;11(20). doi: 10.3390/cells11203190.
6. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22. doi: 10.1093/ije/dyg070.
7. Emdin CA, Khera AV, Kathiresan S. Mendelian Randomization. Jama. 2017;318(19):1925-6. doi: 10.1001/jama.2017.17219.
8. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65. doi: 10.1002/gepi.21758.
9. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14. doi: 10.1002/gepi.21965.
10. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32(5):377-89. doi: 10.1007/s10654-017-0255-x.
11. . Low cigarette consumption and risk of coronary heart disease and stroke: a meta-analysis of 141 cohort studies within 55 study reports. Bmj. 2018;363:k5035. doi: 10.1136/bmj.k5035.
12. Lee SH, Kim BJ, Kang J, Seo DC, Lee SJ. Association of Self-Reported and Cotinine-Verified Smoking Status with Atrial Arrhythmia. J Korean Med Sci. 2020;35(34):e296. doi: 10.3346/jkms.2020.35.e296.
13. Pastori D, Cormaci VM, Marucci S, Franchino G, Del Sole F, Capozza A, et al. A Comprehensive Review of Risk Factors for Venous Thromboembolism: From Epidemiology to Pathophysiology. Int J Mol Sci. 2023;24(4). doi: 10.3390/ijms24043169.
14. Aune D, Schlesinger S, Norat T, Riboli E. Tobacco smoking and the risk of abdominal aortic aneurysm: a systematic review and meta-analysis of prospective studies. Sci Rep. 2018;8(1):14786. doi: 10.1038/s41598-018-32100-2.
15. Zheng L, Baroom G, Naqvi REZ, Heath AK, Berlanga-Taylor A, Hibino M, et al. Tobacco smoking and the risk of aortic aneurysm in the UK biobank. Sci Rep. 2025;15(1):32191. doi: 10.1038/s41598-025-18013-x.
16. Khouja JN, Sanderson E, Wootton RE, Taylor AE, Church BA, Richmond RC, et al. Estimating the health impact of nicotine exposure by dissecting the effects of nicotine versus non-nicotine constituents of tobacco smoke: A multivariable Mendelian randomisation study. PLoS Genet. 2024;20(2):e1011157. doi: 10.1371/journal.pgen.1011157.
17. Saaoud F, Shao Y, Cornwell W, Wang H, Rogers TJ, Yang X. Cigarette Smoke Modulates Inflammation and Immunity via Reactive Oxygen Species-Regulated Trained Immunity and Trained Tolerance Mechanisms. Antioxid Redox Signal. 2023;38(13-15):1041-69. doi: 10.1089/ars.2022.0087.

18. Wei R, Huang M, Zhang Y, He T, Zhu C, Piao H, et al. Tobacco exposure and peripheral artery disease: A multi-omics investigation integrating epidemiology, genetics, toxicology, and single-cell transcriptomics. Ecotoxicology and Environmental Safety. 2026;314:120083. doi: 10.1016/j.ecoenv.2026.120083.
19. Cao S, Zeng Y, Pang K, Chen M, Guo R, Wu N, et al. Unraveling the causal impact of smoking and its DNA methylation signatures on cardiovascular disease: Mendelian randomization and colocalization analysis. Clinical Epigenetics. 2025;17(1):1. doi: 10.1186/s13148-024-01808-6.
20. Li W, Dong P, Li Y, Tang J, Liu S, Tu L, et al. Examining the potential causal relationships among smoking behaviors, blood DNA methylation profiles, and the development of coronary heart disease and myocardial infarction. Clinical Epigenetics. 2024;16(1):173. doi: 10.1186/s13148-024-01791-y.
21. Jones MR, Barnoya J, Stranges S, Losonczy L, Navas-Acien A. Cardiovascular Events Following Smoke-Free Legislations: An Updated Systematic Review and Meta-Analysis. Curr Environ Health Rep. 2014;1(3):239-49. doi: 10.1007/s40572-014-0020-1.
22. Skov-Jeppesen SM, Kobylecki CJ, Jacobsen KK, Bojesen SE. Aryl hydrocarbon receptor repressor (AHRR) methylation predicts risk of vascular disease: A cohort study of the general population. International Journal of Surgery (London, England). 2024;110(11):6953-61. doi: 10.1097/JS9.0000000000001922.
23. Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023;622(7982):329-38. doi: 10.1038/s41586-023-06592-6.
24. Cadiou S, König E, Mapelli A, Pontali G, Ghasemi-Semeskandeh D, Filosi M, et al. Beyond the classical plasma secretome: genetic architecture and disease associations of the expanded human plasma proteome in 13,445 Europeans. 2026:2026.07.23.26358667. doi: 10.64898/2026.07.23.26358667 %J medRxiv.
25. Nikpay M. Joint Analysis of QTL Data Provided Insights into the Connection of Transcriptome and Proteome and the Impact of Omics Platforms. Omics : a Journal of Integrative Biology. 2026;30(10):571-7. doi: 10.1177/15578100261467514.
26. Franchini M, Mannucci PM. ABO blood group and thrombotic vascular disease. Thromb Haemost. 2014;112(6):1103-9. doi: 10.1160/th14-05-0457.
27. Morange P-E, Ibrahim-Kosta M. Ethnicity and Haemostasis: Challenge in the genomics era. Journal of Thrombosis and Haemostasis : JTH. 2020;18(6):1274-5. doi: 10.1111/jth.14864.
28. Manderstedt E, Halldén C, Lind-Halldén C, Elf J, Svensson PJ, Engström G, et al. Thrombotic risk determined by ABO, F8, and VWF variants in a population-based cohort study. Research and Practice in Thrombosis and Haemostasis. 2025;9(4):102875. doi: https://doi.org/10.1016/j.rpth.2025.102875.
29. Tinworth AC, Yao P, Iona A, Pozarickij A, Von Ende A, Millwood IY, et al. Proteome-wide genetic analyses identify ASGR1 as a potential biomarker linking smoking with ischemic heart disease. Nature Cardiovascular Research. 2026;5(10). doi: 10.1038/s44161-026-00871-x.
30. Ishida M, Sakai C, Kobayashi Y, Ishida T. Cigarette Smoking and Atherosclerotic Cardiovascular Disease. Journal of Atherosclerosis and Thrombosis. 2024;31(3):189-200. doi: 10.5551/jat.RV22015.
31. Xiao S, Liu B, Argentieri MA, Belbasis L, Shovlin CL, Collister JA, et al. Proteomic signatures

of smoking and their associations with risk of incident diseases and mortality in diverse populations. Nature Communications. 2025;17(1):928. doi: 10.1038/s41467-025-67656-x.
32. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169(7):1177-86. doi: 10.1016/j.cell.2017.05.038.

Study design

Instrumental SNPs

($P < 5\times10^{-8}$, $r^2 < 0.001$)

Assumption 1

Assumption 2

Assumption 3

Exposures

11 smoking behaviours

Primary:

• Ever smoked (ukb-b-20261)

• Cigarettes per day (GSCAN no-UKB)

• Secondhand smoke (ukb-b-6244)

Sensitivity: 8 additional smoking behaviours

MVMR-adjusted confounders

7 cardiometabolic traits:

T2DM (DIAMANTE no-UKB)

HDL-C, LDL-C, TG (GLGC no-UKB)

Systolic BP (ICBP)

BMI (GIANT), Lp(a) (UKB-PPP)

Outcomes

12 cardiovascular diseases

CAD, HF, PAD, DVT/PE,

AF, DCM, Stroke, VA,

CD, AA, Tachycardia, Bradycardia

Primary: CAD, HF, PAD, DVT/PE

MVMR adjustment

Mechanism

Molecular mechanisms

eQTL-SMR (eQTLGen blood cis-eQTL)

pQTL MR (UKB-PPP Olink)

Bayesian colocalization

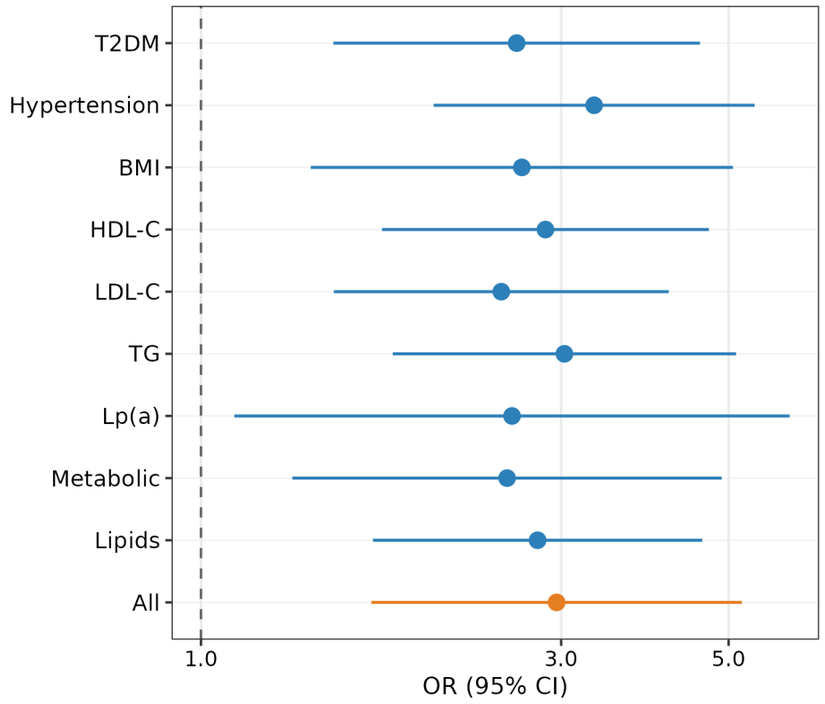
Key findings:

4,278 SMR genes; 18 causal proteins

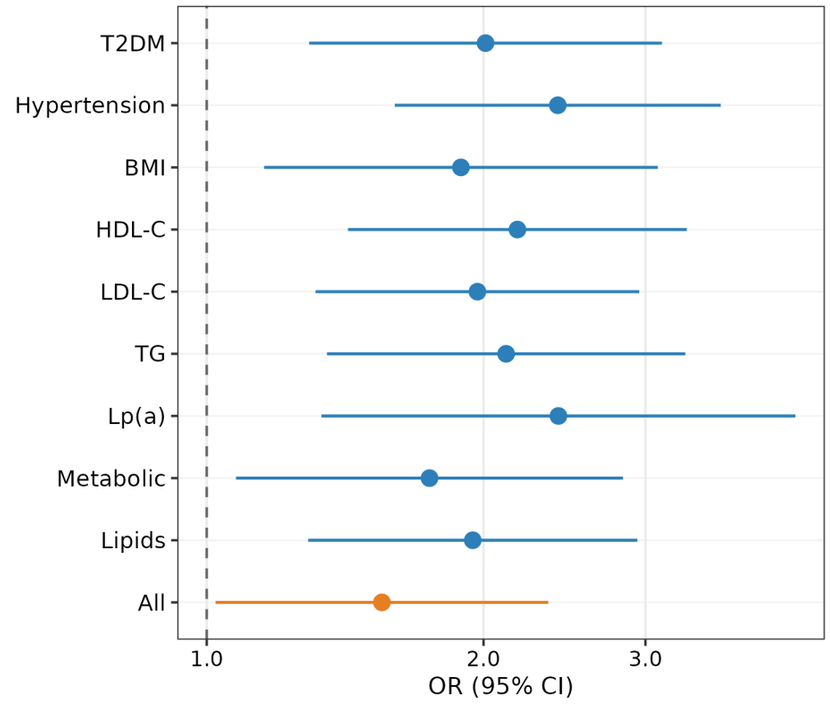
Shared + subtype-specific effectors

PAD: PP.H4 = 0.71 (CHRNA3/5)

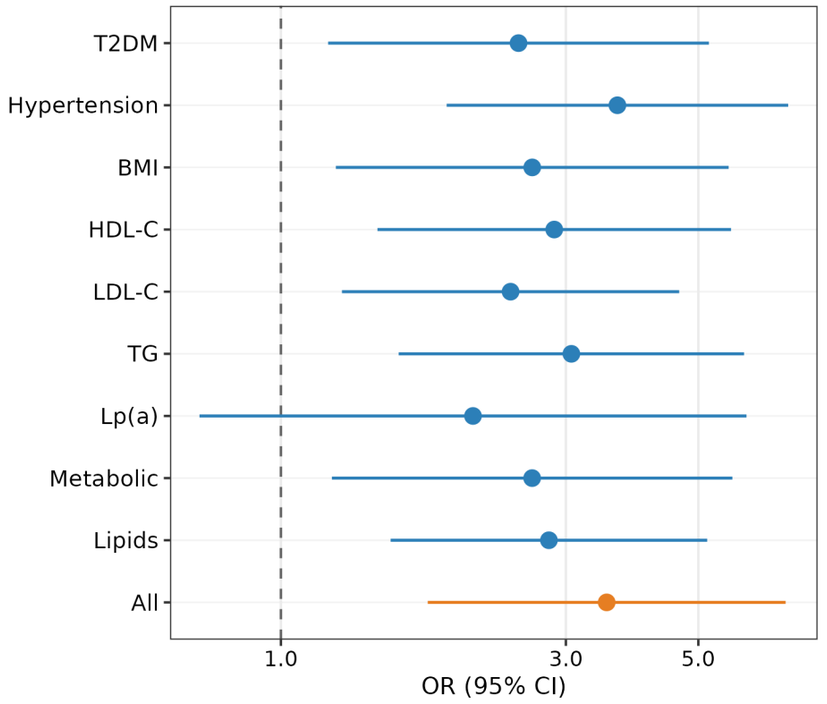
A. Ever smoked → CAD



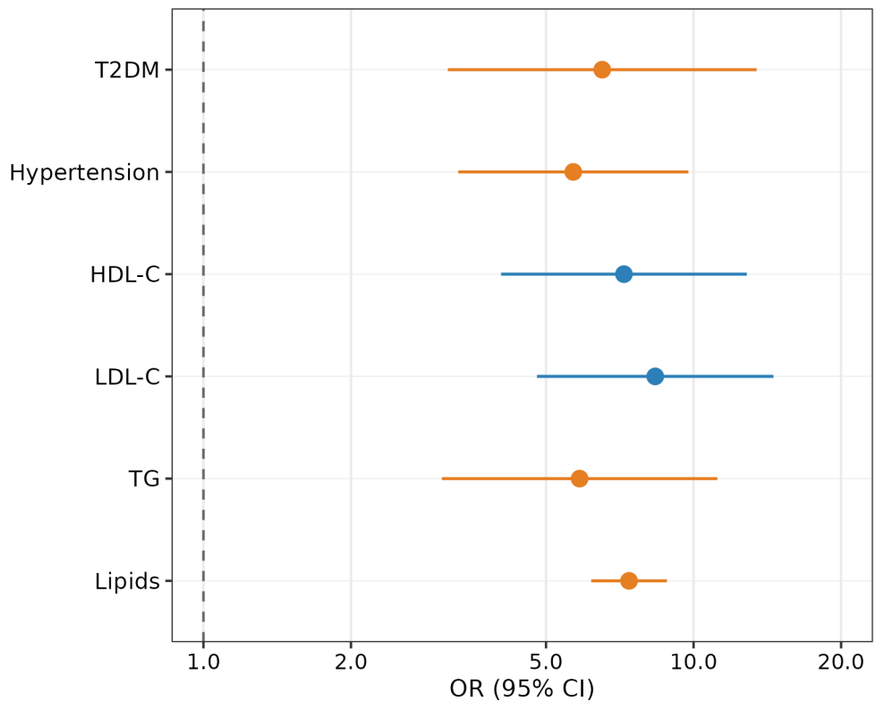
B. Ever smoked → HF



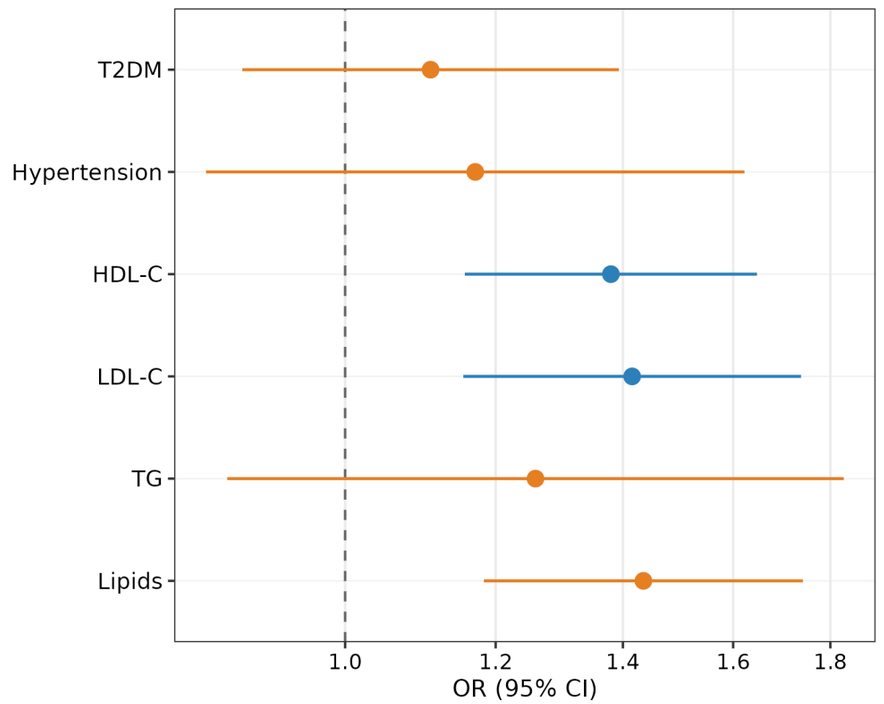
C. Ever smoked → PAD



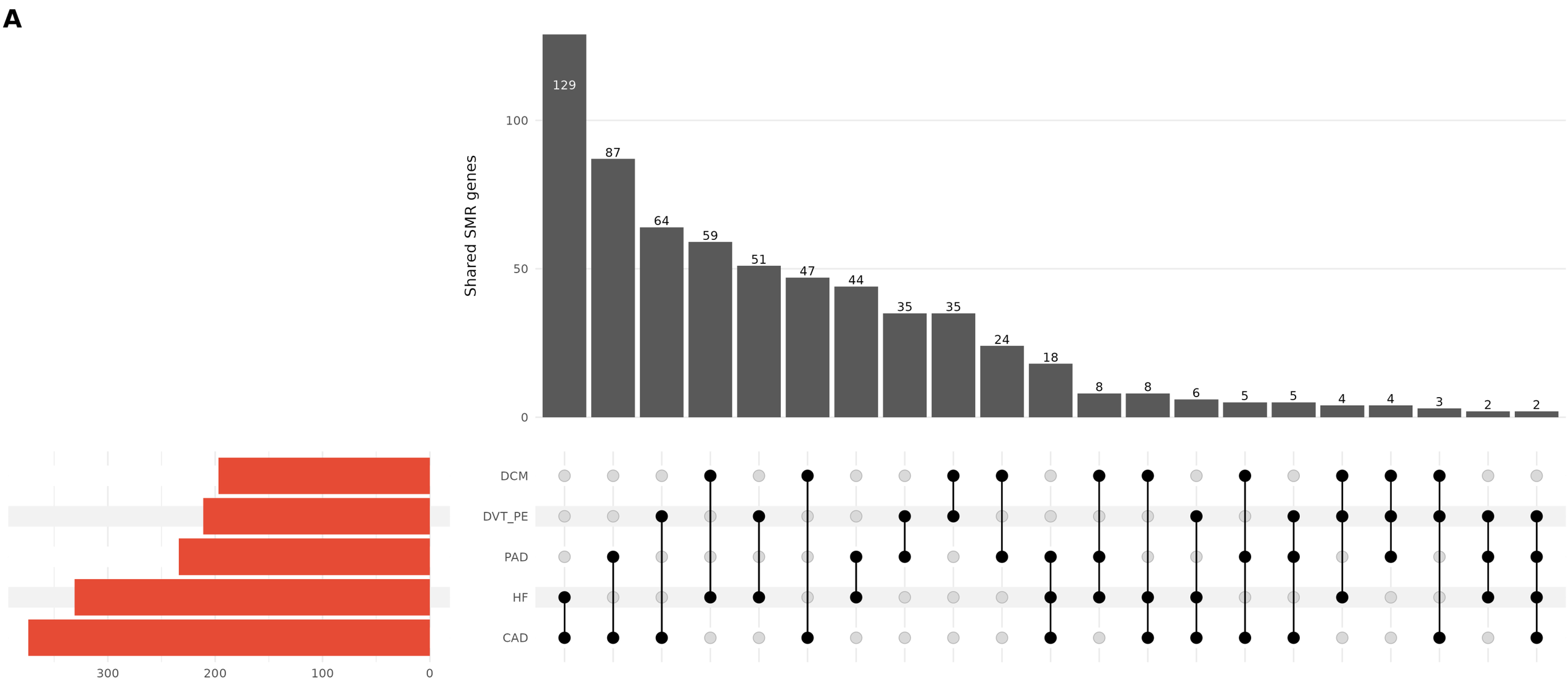
D. Cigarettes per day → PAD



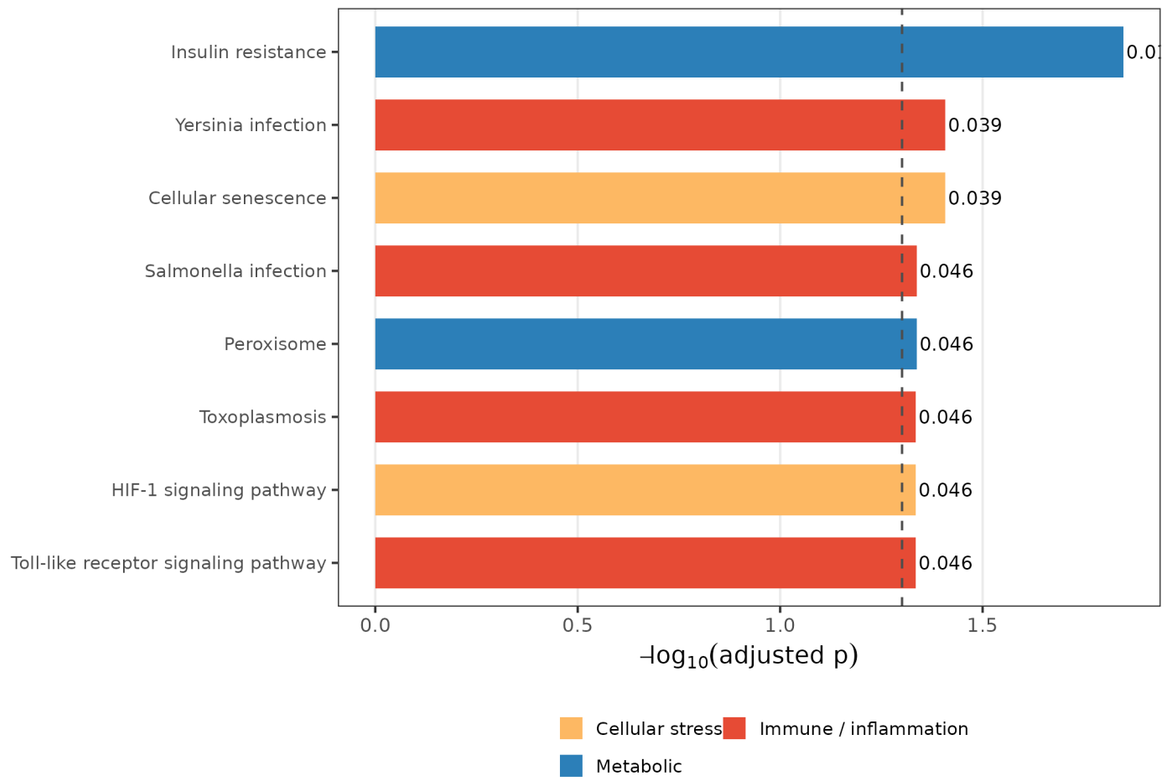
E. Cigarettes per day → HF



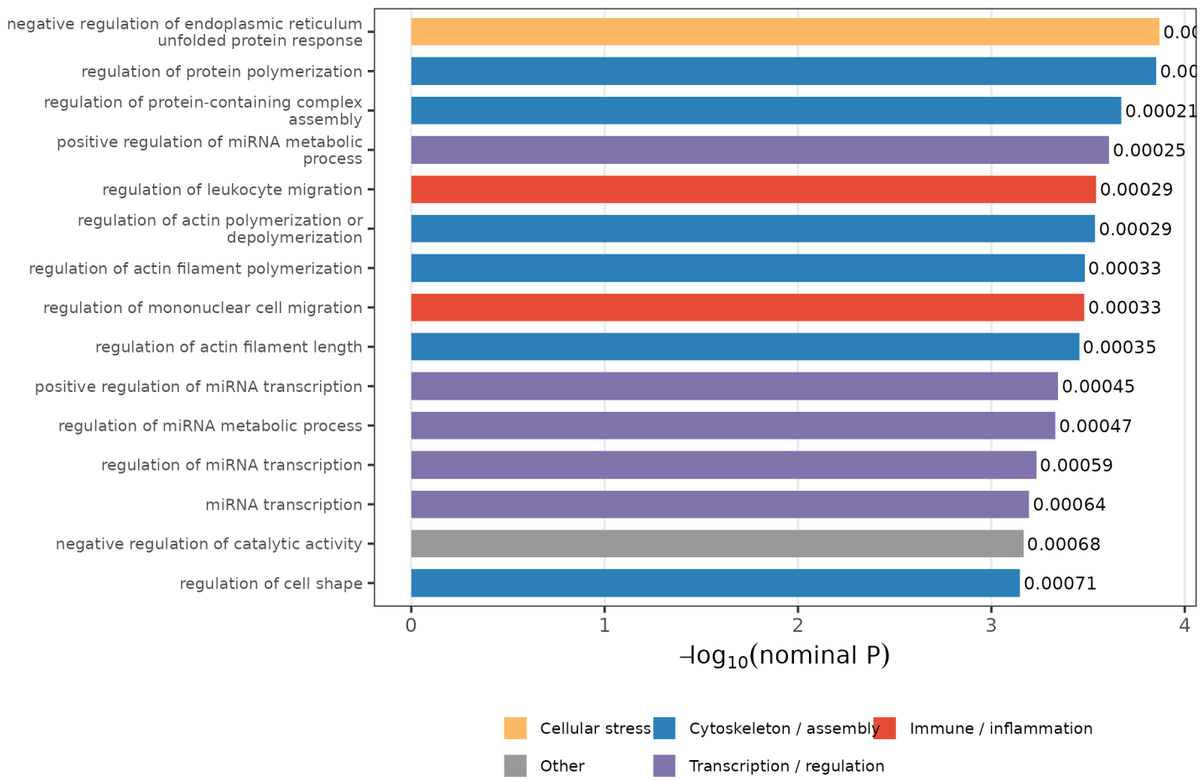
A


Shared SMR genes
129
87
64
59
51
47
44
35
35
24
18
8
8
6
5
5
4
4
3
2
2
100
50
0
DCM
DVT_PE
PAD
HF
CAD
300
200
100
0

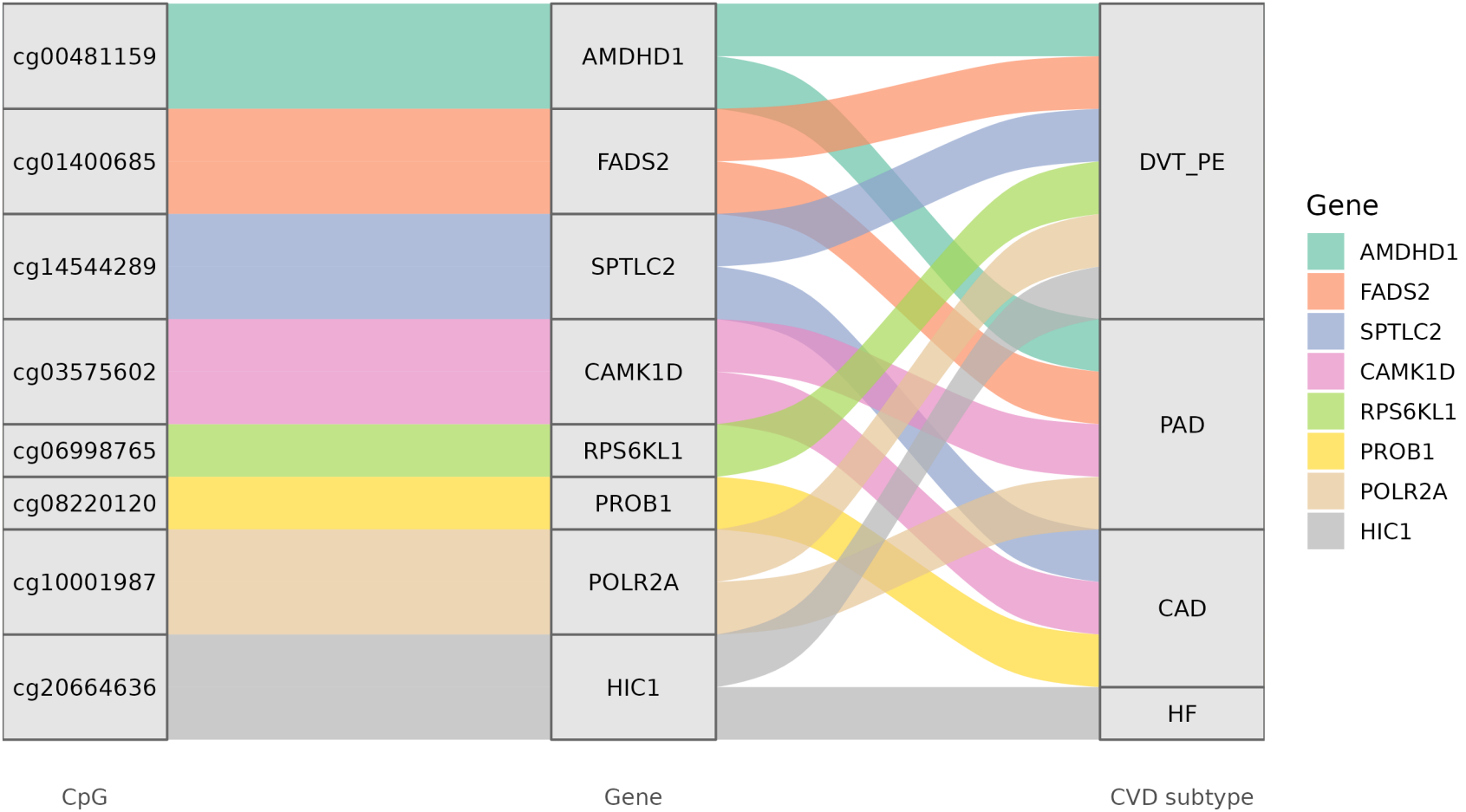
B

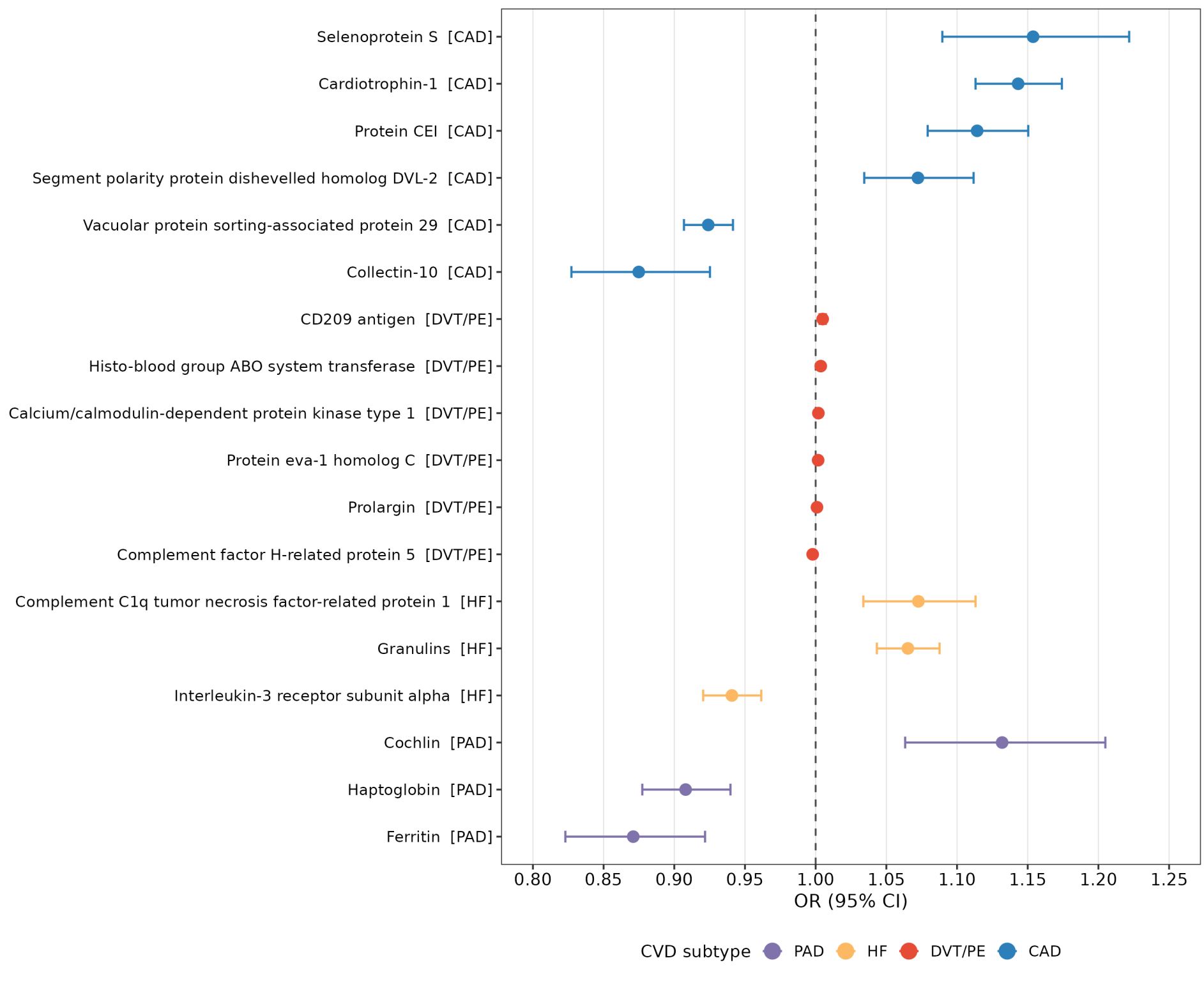

Insulin resistance
Yersinia infection
Cellular senescence
Salmonella infection
Peroxisome
Toxoplasmosis
HIF-1 signaling pathway
Toll-like receptor signaling pathway
0.039
0.039
0.046
0.046
0.046
0.046
0.046
0.0
0.5
1.0
1.5
−log10(adjusted p)
Cellular stress
Immune / inflammation
Metabolic

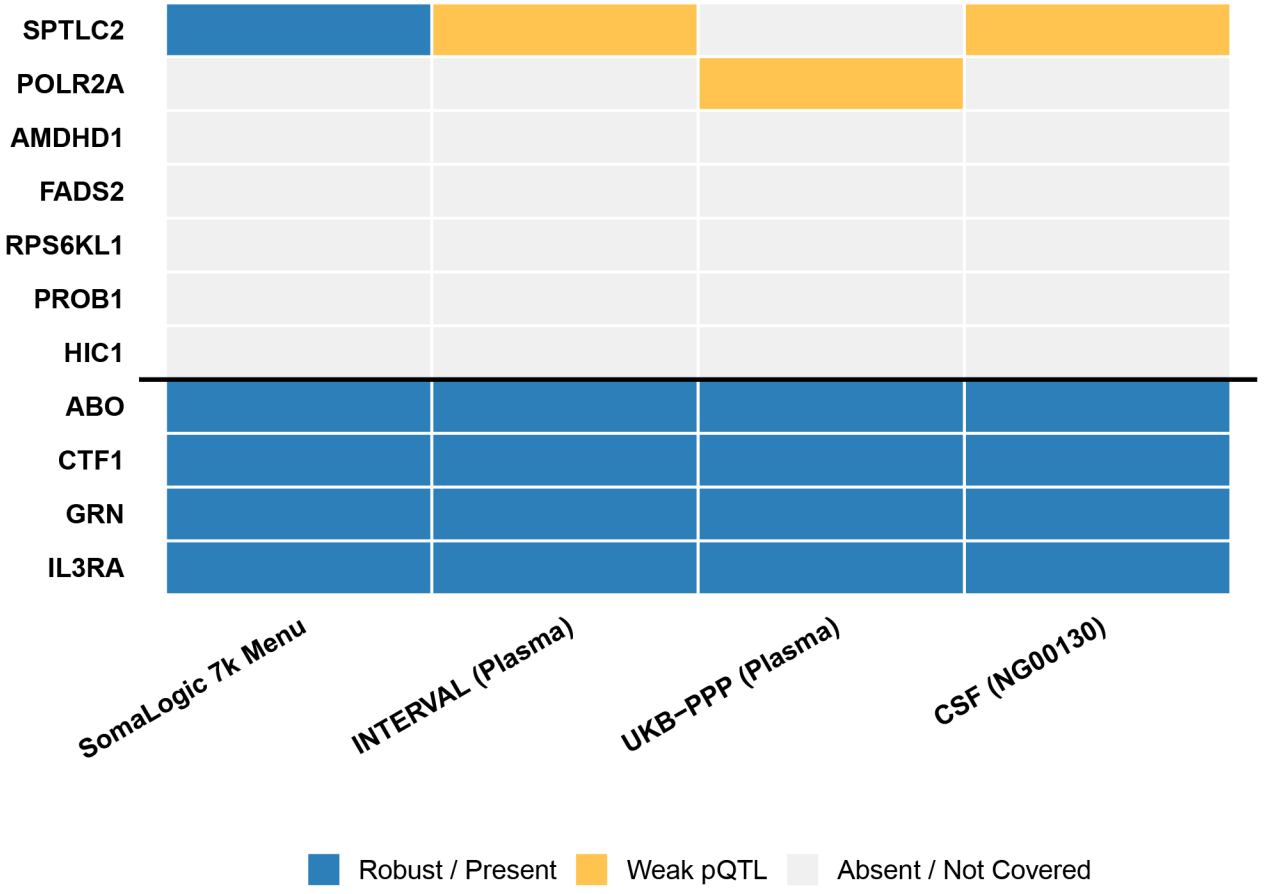
C

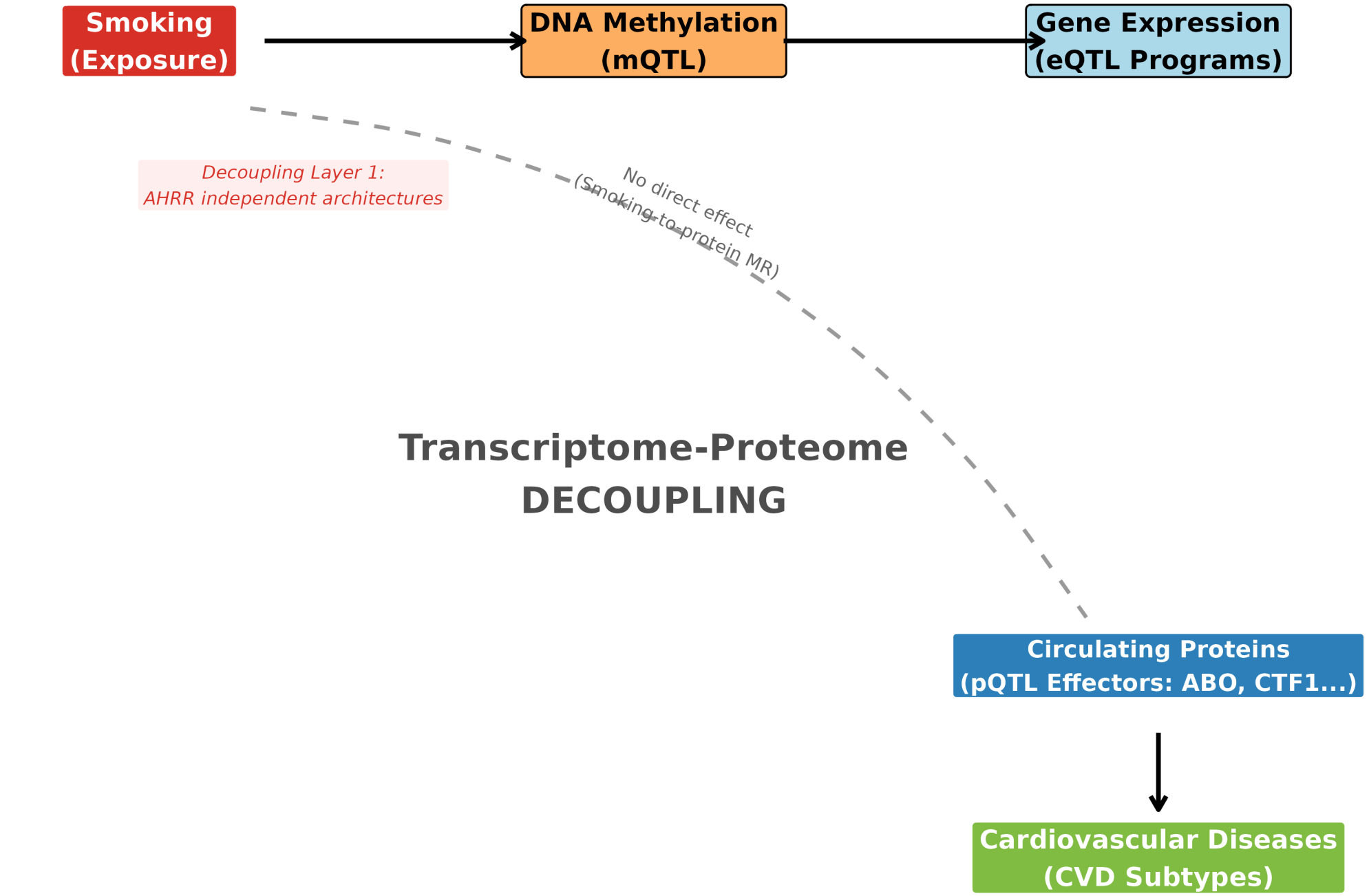

negative regulation of endoplasmic reticulum unfolded protein response
regulation of protein polymerization
regulation of protein-containing complex assembly
positive regulation of miRNA metabolic process
regulation of leukocyte migration
regulation of actin polymerization or depolymerization
regulation of actin filament polymerization
regulation of mononuclear cell migration
regulation of actin filament length
positive regulation of miRNA transcription
regulation of miRNA metabolic process
regulation of miRNA transcription
miRNA transcription
negative regulation of catalytic activity
regulation of cell shape
0.00021
0.00025
0.00029
0.00029
0.00033
0.00033
0.00035
0.00045
0.00047
0.00059
0.00064
0.00068
0.00071
0
1
2
3
4
−log10(nominal P)
Cellular stress
Cytoskeleton / assembly
Immune / inflammation
Other
Transcription / regulation

Smoking-associated CpGs → shared eQTL-SMR genes → CVD subtypes




Selenoprotein S [CAD]
Cardiotrophin-1 [CAD]
Protein CEI [CAD]
Segment polarity protein dishevelled homolog DVL-2 [CAD]
Vacuolar protein sorting-associated protein 29 [CAD]
Collectin-10 [CAD]
CD209 antigen [DVT/PE]
Histo-blood group ABO system transferase [DVT/PE]
Calcium/calmodulin-dependent protein kinase type 1 [DVT/PE]
Protein eva-1 homolog C [DVT/PE]
Prolargin [DVT/PE]
Complement factor H-related protein 5 [DVT/PE]
Complement C1q tumor necrosis factor-related protein 1 [HF]
Granulins [HF]
Interleukin-3 receptor subunit alpha [HF]
Cochlin [PAD]
Haptoglobin [PAD]
Ferritin [PAD]
0.80
0.85
0.90
0.95
1.00
1.05
1.10
1.15
1.20
1.25
OR (95% CI)
CVD subtype
PAD
HF
DVT/PE
CAD


SPTLC2
POLR2A
AMDHD1
FADS2
RPS6KL1
PROB1
HIC1
ABO
CTF1
GRN
IL3RA
SomaLogic 7k Menu
INTERVAL (Plasma)
UKB–PPP (Plasma)
CSF (NG00130)
Robust / Present
Weak pQTL
Absent / Not Covered


Smoking (Exposure)
DNA Methylation (mQTL)
Gene Expression (eQTL Programs)
Decoupling Layer 1: AHRR independent architectures
No direct effect (Smoking-to-protein MR)
Transcriptome-Proteome DECOUPLING
Circulating Proteins (pQTL Effectors: ABO, CTF1...)
Cardiovascular Diseases (CVD Subtypes)
Pathway-Level Convergence: Metabolic, Inflammatory, Cellular Stress, Complement

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigureS2.tif
- SupplementaryFigureS1.tif
- SupplementaryFigureS3.tif
- SupplementaryTable.xlsx