

1 Genetic correlation and causality between smoking and 42 2 neuropsychiatric and gastrointestinal diseases

3 Jiayi Xiong¹, Zheng Wang^{1,3,4}, Yuanfeng Huang¹, Shiyu Zhang^{1,5}, Guang Yang⁷, Jiaqi Yang¹, Shuo
4 Gao¹, Tianyang Wang^{1,5}, Jinchen Li^{1,2,3,6*}, Guihu Zhao^{1,3*}, Bin Li^{1,3*}

5
6
7
8 *Correspondence:

9 Jinchen Li, Guihu Zhao, Bin Li

10 Address: No.87 Xiangya Road, Changsha, Hunan, China

11 Telephone: 0731-89752068

12 E-mail address: lijinchen@csu.edu.cn, ghzhao@csu.edu.cn, lebin001@csu.edu.cn

13

14

15

16 ¹Bioinformatics Centre, National Clinical Research Centre for Geriatric Disorders, Department of
17 Geriatrics, Xiangya Hospital, Central South University, Changsha, Hunan 410008, China.

18 ²Centre for Medical Genetics & Hunan Key Laboratory, School of Life Sciences, Central South
19 University, Changsha, Hunan 410008, China.

20 ³Department of Neurology, Xiangya Hospital, Central South University, Changsha, Hunan 410008,
21 China.

22 ⁴Hunan Key Laboratory of Molecular Precision Medicine, Xiangya Hospital, Central South University,
23 Changsha, Hunan 410008, China.

24 ⁵Xiangya School of Medicine, Central South University, Changsha, Hunan 410013, China.

25 ⁶Bioinformatics Centre, Furong Laboratory, Changsha, Hunan 410008, China.

26 ⁷Department of Orthopedics, Xiangya Hospital, Central South University, Changsha 410008, Hunan,
27 China.

28

29

30

31 Jiayi Xiong: xiongjy@csu.edu.cn

32 Zheng Wang: wangzheng888@csu.edu.cn

33 Yuanfeng Huang: 2025512649@qq.com

34 Shiyu Zhang: shiyuzhang0522@gmail.com

35 Guang Yang: 228111092@csu.edu.cn

36 Jiaqi Yang: 228111036@csu.edu.cn

37 Shuo Gao: 228111037@csu.edu.cn

38 Tianyang Wang: wangtianyang@csu.edu.cn

39

40

41

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

42 **Abstract**

43 **Background**

44 Previous investigations have elucidated epidemiological associations linking smoking to
45 neuropsychiatric and gastrointestinal diseases, yet the underlying causal relationships remain
46 enigmatic. To shed light on this matter, we undertook a Mendelian randomisation(MR) study
47 with the aim of gauging the potential causal association between smoking and the susceptibility
48 to neuropsychiatric and gastrointestinal tract diseases.

49 **Methods**

50 We meticulously collected and preprocessed genome-wide association study (GWAS) data
51 encompassing smoking (280,508 cases and 180,558 controls) as well as neuropsychiatric and
52 gastrointestinal phenotypes (n= 6,681 to 87,3341). To investigate the genetic correlation
53 between smoking and diseases, we employed linkage disequilibrium score regression. We
54 further applied multi-trait analysis of GWAS to identify the shared risk single-nucleotide
55 polymorphisms (SNPs) implicated in both smoking and diseases. Pleiotropic genes were
56 annotated by enrichment analysis. Subsequently, bidirectional MR analysis was performed to
57 infer causality.

58 **Results**

59 Our findings, supported by robust evidence derived from an expansive sample size,
60 demonstrate that smoking exerts a causal influence on merely six of these diseases, while no
61 disease was found to causally impact smoking. Intriguingly, we discovered 513 pleiotropic
62 genes enriched in pathways such as the regulation of growth and synapses, suggesting a

63 potential shared genetic basis between smoking and these diseases, leading to aberrant neural
64 development. Remarkably, among the 42 diseases scrutinized, a significant genetic correlation
65 was exclusively observed with gastroesophageal reflux disease (GRED). Furthermore, we
66 identified risk SNPs shared by smoking and GRED.

67 Conclusions

68 This study revealed the shared genetic basis and causal effects connecting smoking to
69 neuropsychiatric and gastrointestinal diseases, thereby providing novel etiological insights into
70 the role of smoking in these diseases.

71 Keywords

72 Smoking; Neuropsychiatric diseases; Gastrointestinal diseases; Mendelian randomisation;
73 Genetic correlation

74 Background

75 Tobacco utilization constitutes a momentous global public health concern and stands as a
76 paramount contributor to preventable mortality(1). Previous epidemiological inquiries have
77 surfaced, underscoring smoking as a risk factor for many diseases, including cancers,
78 cardiovascular diseases, chronic obstructive pulmonary disease and so on(2). In recent years,
79 the advent of genome-wide association studies (GWAS) has engendered a burgeoning body of
80 evidence, illuminating the shared genetic underpinnings between smoking and various diseases.
81 Genetic correlation analysis, predicated on GWAS summary statistics, has unveiled the genetic
82 intertwinement linking smoking to metabolic disorders, autoimmune conditions, psychiatric
83 disorders and nervous system diseases(3–5). Beyond genetic correlation, methodologies such

84 as Mendelian randomisation (MR), which uses genetic variants as instrumental variables, have
85 been employed to discern the causal connections between lifestyle choices and diseases(6).
86 Previous MR endeavors have provided insights into the causal ramifications of smoking on
87 diseases such as lung cancer, coronary heart disease, and type 2 diabetes(7–9).
88 It is firmly established that smoking bears an association with neuropsychiatric and
89 gastrointestinal diseases, both of which bear an intimate nexus with the brain-gut axis(10,11).
90 The brain-gut axis epitomizes the bidirectional communication interlinking the central nervous
91 system and the gastrointestinal tract, constituting a pivotal arbiter for homeostasis(10).
92 Perturbations within the purview of the brain-gut axis have been inextricably implicated in an
93 assortment of ailments. For instance, patients afflicted with irritable bowel syndrome have
94 evinced perturbed interactions within the brain-gut axis(12). Although smoking has been linked
95 to both types of diseases, including Parkinson’s disease(13) and colorectal cancer(14), in
96 epidemiological studies, the extent of genetic correlation and causation between smoking and
97 these diseases remains largely unknown.

98 This pioneering study represents an all-encompassing inquiry, scrutinizing the genetic
99 correlation between smoking and 42 neuropsychiatric and gastrointestinal diseases.
100 Furthermore, we unveiled the shared genetic substratum and the potential causative associations.
101 Our approach embraces an expansive and comprehensive sample size, encompassing a wide
102 spectrum of diseases, thereby fomenting a holistic comprehension of the mechanistic role
103 played by smoking. This distinguishing facet sets our investigation apart from previous research,
104 emphasising the profound genetic insights it yields.

105 **Methods**

106 **2.1 Study overview and data resources**

107 Briefly, this study obtained GWAS summary statistics for smoking and 42 neuropsychiatric
108 and gastrointestinal diseases. The GWAS summary statistics for smoking and the diseases were
109 sourced from publicly available databases. Specifically, the smoking GWAS data originated
110 from a meta-analysis encompassing the UK Biobank and the Tobacco and Genetics Consortium
111 datasets, focusing on smoking status (ever versus never smoked). This dataset included 280,508
112 individuals who were ever smokers and 180,558 individuals who were never smokers, all of
113 European ancestry. The data were obtained from the esteemed MRC-IEU OpenGWAS database,
114 with the accession ID: ukb-b-20261 (accessible at <https://gwas.mrcieu.ac.uk/>). The GWAS data
115 pertaining to the 42 diseases were acquired from the MRC-IEU OpenGWAS database, the
116 Psychiatric Genomics Consortium (PGC), and the GWAS catalog, ensuring a sample size
117 exceeding 5,000. Detailed summary statistics for the GWAS data employed in this study can
118 be found in Table 1.

119

120

Table 1 | Details of GWAS Summary Data Sources

Disease group	Diseases	PMID/ IEUID	N_Cases	N_Controls	N_total	Journal	Source
Psychiatric diseases	ADHD	30478444	19099	34194	53293	Nature Genetics	PGC
	Anorexia nervosa	31308545	16992	55525	72517	Nature Genetics	PGC
	BD	34002096	41917	371549	413466	Nature Genetics	PGC
	Cannabis use disorder	33096046	14080	343726	357806	Lancet Psychiatry	PGC
	Panic disorder	34737426	2248	7992	10240	Mol Psychiatry	PGC
	MDD	30718901	170756	329443	500199	Nature Neurosci	PGC
	OCD (Tourette syndrome)	30818990	4819	9488	14307	Am J Psych	PGC
	PTSD	31594949	23212	151447	174659	Nature Commun	PGC

	Schizophrenia	35396580	52017	75889	127906	Nature	PGC
	ASD	32747698	22196	32504	55420	Transl Psychiatry	GWAS Catalog
	MS	3160424	47429	68374	115803	Science	MRC-IEU
	ALS	34873335	27205	110881	138086	Nat Genet	GWAS Catalog
Nervous system diseases	AD	35379992	85934	401577	487511	Nat Genet	GWAS Catalog
	Lewy body dementia	33589841	2591	4027	6681	Nat Genet	GWAS Catalog
	Epilepsy	ukb-b-1600	2326	460684	463010	-	MRC-IEU
	Migraine	35115687	102084	771257	873341	Nat Genet	GWAS Catalog
	Stroke	34737426	6986	448317	455303	Nat Genet	GWAS Catalog
	PD	ieu-b-7	33674	449056	482730	Nat Genet	MRC-IEU
	Dysphagia	34737426	4601	451747	456348	Nat Genet	GWAS Catalog
	GRED	34187846	129080	473524	602604	Gut	GWAS Catalog
Gastrointestinal diseases	Gastritis	34737426	17420	438928	456348	Nat Genet	GWAS Catalog
	GU	34737426	2413	453935	456348	Nat Genet	GWAS Catalog
	Dyspepsia	ukb-b-14814	7662	455348	463010	-	MRC-IEU
	IBS	34741163	53400	433201	486601	Nat Genet	GWAS Catalog
	CD	34737426	1342	455006	456348	Nat Genet	GWAS Catalog
	UC	34737426	2569	453779	456348	Nat Genet	GWAS Catalog
	IBD	33959723	4101	480497	484598	Nat Aging	GWAS Catalog
	Colon cancer	34737426	1482	454794	456276	Nat Genet	GWAS Catalog
	Diverticular disease	30661054	31964	419135	451099	Gut	GWAS Catalog
	Duodenitis	34737426	2286	454062	456348	Nat Genet	GWAS Catalog
	Constipation	34737426	4781	451567	456348	Nat Genet	GWAS Catalog
	Fatty liver disease	34841290	8434	770180	778614	Cell Rep Med	GWAS Catalog
	Cholelithiasis	34737426	7426	448922	456348	Nat Genet	GWAS Catalog
	Acute pancreatitis	34737426	1748	454600	456348	Nature	GWAS Catalog
	Acute appendicitis	34737426	2588	453760	456348	Nat Genet	GWAS Catalog
	Peritonitis	ukb-b-11751	1555	461378	462933	-	MRC-IEU
	DU	ukb-d-K26	1291	359903	361194	-	MRC-IEU
	Polyp of colon	ukb-b-17845	2437	460573	463010	-	MRC-IEU
Cholecystitis	ukb-d-K81	1930	359264	361194	-	MRC-IEU	
Inguinal hernia	ukb-a-549	8815	328384	337199	-	MRC-IEU	
Oesophagitis	ukb-b-19354	4931	458079	463010	-	MRC-IEU	
Primary biliary cholangitis	34033851	8021	16489	24510	-	MRC-IEU	

Abbreviations: PMID, PubMed unique identifier; IEUID, the ID of the GWAS dataset in MRC IEU OpenGWAS; ADHD, attention deficit hyperactivity disorder; BD, bipolar disorder; MDD, major depressive disorder; OCD, obsessive-compulsive disorder; PTSD, post-traumatic stress disorder; ASD, autism spectrum disorders; MS, multiple sclerosis; ALS, amyotrophic lateral sclerosis; AD, Alzheimer's disease; PD, Parkinson's disease; GRED, gastroesophageal reflux disease; GU, gastric ulcer; IBS, irritable bowel syndrome; CD, Crohn's disease; UC, ulcerative colitis; IBD, inflammatory bowel disease; DU, duodenal ulcer.

121

122

123

124 **2.2 Statistical Analysis**

125 **2.2.1 Genetic correlation analysis**

126 To explore the genetic correlation between smoking and the aforementioned 42 diseases, we
127 employed linkage disequilibrium score regression (LDSC)(15). This method involves
128 regressing the product of z scores from two GWAS scans against a measure of linkage
129 disequilibrium for each single nucleotide polymorphism (SNP). The genetic correlation results
130 were subjected to correction for multiple testing using the rigorous Bonferroni correction,
131 ensuring the statistical integrity of these findings. Genetic correlations with a P-value below the
132 significance threshold ($0.05/42 = 1.25 \times 10^{-3}$) were deemed statistically significant.

133 **2.2.2 Pleiotropy analysis**

134 For diseases exhibiting significant genetic correlation with smoking according to LDSC
135 analysis, we conducted multitrait analysis of GWAS (MTAG)(16,17) to identify shared
136 pleiotropic SNPs between smoking and the respective diseases. MTAG is an advanced
137 statistical fine-mapping method that leverages summary statistics from multiple GWAS to
138 enhance the power of detecting SNPs associated with traits. It models the effect sizes of each
139 SNP on different traits, assuming a distribution with a variance–covariance matrix proportional
140 to the genetic correlation matrix. MTAG was executed with default parameters, and the
141 genome-wide significance threshold was set at $P < 5 \times 10^{-8}$. The results include a residual
142 correlation matrix between estimated traits, a genetic correlation matrix, and association
143 statistics for each SNP.

144 To discern pleiotropic genes, we employed MAGMA(18), a gene and pathway analysis tool

145 grounded in a multiple linear principal components regression model utilizing GWAS summary
146 statistics. The reference panel employed was the linkage disequilibrium structure derived from
147 European ancestry samples of the 1000 Genomes Project Phase 3. Genes exhibiting a false
148 discovery rate (FDR)-corrected P value below 0.05 in both smoking and at least one disease
149 were defined as pleiotropic genes.

150 Pathway enrichment analysis of the identified pleiotropic genes was performed using
151 Metascape(19). The enriched Gene Ontology (GO) terms were visually represented using the
152 clusterProfiler(20) R package.

153 **2.2.3 Causal inference analysis**

154 To infer potential causal relationships between smoking and the diseases under investigation,
155 we conducted bidirectional MR analysis using the R package TwoSampleMR(21). The GWAS
156 summary statistics for smoking served as the exposure variable, while each disease was
157 considered an outcome variable. Five different MR methods, including inverse variance
158 weighted (IVW), MR-Egger, weighted median, simple mode, and weighted mode methods,
159 were adopted. The IVW method was selected as the primary approach due to its high efficiency
160 and minimal assumptions(22). SNPs with an F statistic greater than 10 were included as
161 instrumental variables. The MR results were adjusted for multiple testing using the FDR
162 method. Causal associations were reported solely for analysis without significant heterogeneity
163 or horizontal pleiotropy, as determined by both Cochran's Q test and the MR-PRESSO test.
164 For a significant causal relationship, we conducted a leave-one-out analysis to assess whether
165 the results were driven by any individual SNP.

166 **Results**

167 **3.1 Summary information for GWAS data**

168 The GWAS summary statistics used in this study are presented in Table 1. The smoking GWAS
169 data encompassed 280,508 individuals classified as ever smokers and 180,558 individuals
170 classified as never smokers of European ancestry, acquired from the MRC-IEU OpenGWAS
171 database (ID: ukb-b-20261). The GWAS data for the 42 neuropsychiatric and gastrointestinal
172 diseases were derived from individuals of European ancestry with sample sizes greater than
173 5,000 (n = 6,681 to 87,3341). (Table 1)

174 **3.2 Genetic correlation analysis**

175 The genetic correlation results between smoking and these 42 diseases estimated by LDSC are
176 shown in Table 2. Following Bonferroni correction, only one disease, gastroesophageal reflux
177 disease (GRED), exhibited a statistically significant negative genetic correlation with smoking
178 ($r_g = -0.226$, $P = 1.27 \times 10^{-7}$). The negative genetic correlation suggests the existence of shared
179 genetic factors between smoking and GRED, albeit with an inverse association. This implies
180 that certain genetic variants associated with smoking may confer a protective effect against the
181 development of GRED. This discovery holds valuable implications for our understanding of
182 the genetic determinants underlying smoking behavior and its potential impact on disease
183 susceptibility. Further investigation is warranted to unravel the precise mechanisms underlying
184 this relationship and explore potential therapeutic avenues for GRED.

185

Table 2 | Genetic Correlation Between “ever smoked” and 42 Diseases Estimated by LDSC

Disease group	Diseases	r_g (se)	P	Intercept (se)
---------------	----------	------------	---	----------------

	ADHD	-0.0294(0.0196)	0.1332	0.0143(0.0073)
	Anorexia nervosa	-0.083(0.0616)	0.178	0.0149(0.0173)
	BD	0.0474 (0.027)	0.0791	0.0205 (0.0126)
	Cannabis use disorder	0.0486(0.0255)	0.0564	-0.0044(0.0072)
Psychiatric diseases	Panic disorder	0.0114(0.0359)	0.7501	0.0006(0.0067)
	MDD	-0.022(0.0156)	0.1598	0.0236(0.0084)
	OCD (Tourette syndrome)	0.0136(0.0247)	0.5825	-0.0101(0.0069)
	PTSD	-0.018(0.0402)	0.6539	0.0022(0.0069)
	Schizophrenia	-0.1024(0.038)	0.007	-0.0379(0.0175)
	ASD	-0.028(0.0181)	0.1206	0.0053(0.0069)
	MS	-0.0032(0.0412)	0.938	-0.0122(0.0139)
Nervous system diseases	ALS	0.0166(0.0206)	0.4207	-0.0085(0.0072)
	AD	-0.0204(0.0176)	0.246	0.0117(0.0076)
	Lewy body dementia	-0.0392(0.0957)	0.6824	-0.0035(0.0101)
	Epilepsy	0.0926(0.1335)	0.4883	0.014(0.0155)
	Migraine	0.0388(0.0365)	0.2877	0.0023(0.0129)
	Stroke	0.0573(0.0674)	0.3953	0.014(0.0097)
	PD	-0.0166(0.0531)	0.7546	-0.013(0.014)
	Dysphagia	0.1192(0.0721)	0.0984	-0.0058(0.0095)
	GREED	-0.2259(0.0428)	1.2731E-07*	-0.0724(0.0179)
	Gastritis	0.0197(0.0565)	0.7277	0.0152(0.0101)
	GU	0.0242(0.1006)	0.8097	-0.0018(0.0095)
	Dyspepsia	-	-	-
	IBS	-0.0267(0.0193)	0.1666	0.0114(0.0081)
	CD	-0.0353(0.0636)	0.5795	0.0123(0.0104)
	UC	0.1042(0.0473)	0.0276	-0.0064(0.0104)
	IBD	0.0342(0.0739)	0.643	-0.0284(0.0136)
	Colon cancer	0.0416(0.0997)	0.6762	0.0037(0.0097)
Gastrointestinal diseases	Diverticular disease	0.0009(0.0445)	0.9831	-0.0604(0.0158)
	Duodenitis	-0.0966(0.0811)	0.2336	0.0211(0.0092)
	Constipation	0.0072(0.1639)	0.965	0.0106(0.0113)
	Fatty liver disease	-0.0457(0.0353)	0.1954	0.0087(0.007)
	Cholelithiasis	-0.0084(0.0513)	0.8703	0.0033(0.0103)
	Acute pancreatitis	-0.0138(0.1031)	0.8933	0.0065(0.0105)
	Acute appendicitis	-0.1265(0.0834)	0.1292	0.0186(0.0092)
	Peritonitis	0.6043(0.9047)	0.5041	-0.0176(0.0204)
	DU	0.1532(0.1353)	0.2574	0.003(0.013)
	Polyp of colon	-	-	-0.0002(0.0147)
	Cholecystitis	-0.1388(0.1194)	0.2451	0.0272(0.0129)
	Inguinal hernia	0.004(0.0682)	0.9538	0.0123(0.0133)
	Oesophagitis	0.218(0.1189)	0.0668	0.0053(0.0136)
	Primary biliary cholangitis	0.0332(0.0519)	0.5231	0.0203(0.0218)

Abbreviations: LDSC, linkage disequilibrium score regression; rg, genetic correlation estimate; se, standard error; ADHD, attention deficit hyperactivity disorder; BD, bipolar disorder; MDD, major depressive disorder; OCD, obsessive-compulsive disorder; PTSD, post-traumatic stress disorder; ASD, autism spectrum disorders; MS, multiple sclerosis; ALS, amyotrophic lateral sclerosis; AD, Alzheimer's disease; PD, Parkinson's disease; GRED, gastroesophageal reflux disease; GU, gastric ulcer; IBS, irritable bowel syndrome; CD, Crohn's disease; UC, ulcerative colitis; IBD, inflammatory bowel disease; DU, duodenal ulcer.

*P value reached the Bonferroni-corrected significance level $P < 1.25 \times 10^{-3}$ (0.05/42)

186

187 **3.2 Pleiotropy analysis**

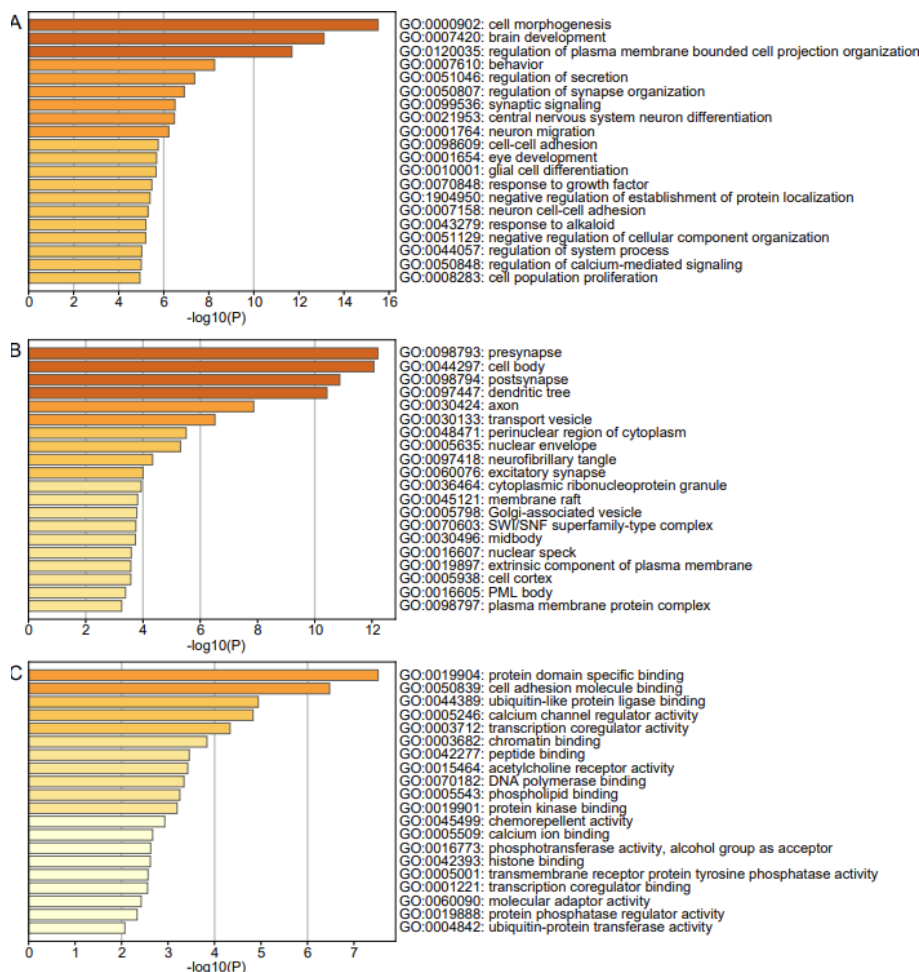
188 Given that only GRED showed a significant genetic correlation with smoking, we proceeded
189 to perform MTAG between the GWAS data for smoking and GRED. A total of 971,487 SNPs
190 were included in the meta-analysis by MTAG. Among them, 358 SNPs reached the genome-
191 wide significance level ($P < 5 \times 10^{-8}$) in the MTAG smoking analysis, indicating their status as
192 pleiotropic SNPs shared by smoking and GRED (detailed information provided in Additional
193 file: Supplementary Table S1).

194 Using MAGMA, we identified 513 pleiotropic genes associated with both smoking and at least
195 one disease after excluding genes on chromosome 6 (Additional file: Supplementary Table S2).

196 These genes are considered candidate pleiotropic genes that are shared between smoking and
197 multiple diseases, indicating their potential role in modulating the relationship between
198 smoking and these diseases. Notably, genes such as CRHR1, KANSL1, PLEKHM1, STH and
199 WNT3 on chromosome 17 as well as MSRA on chromosome 8 were identified as pleiotropic
200 genes in at least 5 diseases respectively.

201 Pathway enrichment analysis of the identified pleiotropic genes revealed significant enrichment
202 in multiple GO terms (Figure 1; Additional file: Supplementary Table S3-S5), encompassing
203 processes such as cell projection morphogenesis, presynapse, and protein domain specific
204 binding. Importantly, our findings indicate that smoking interacts with these signaling pathways,

205 contributing to the occurrence of 42 neuropsychiatric and gastrointestinal diseases. Remarkably,
 206 our results align with previous literature, substantiating the consistent relationship between
 207 smoking and these diseases at a molecular level.



208
 209 **Fig. 1 Enrichment maps of pleiotropic genes for smoking and 42 diseases.** A GO biological
 210 process enrichment bar chart. B GO cellular component enrichment bar chart. C GO molecular function
 211 enrichment bar chart. From panels A to C, the X-axis is $-\log_{10}(P)$, and the Y-axis is the GO term. The
 212 colour represents the enriched significance.

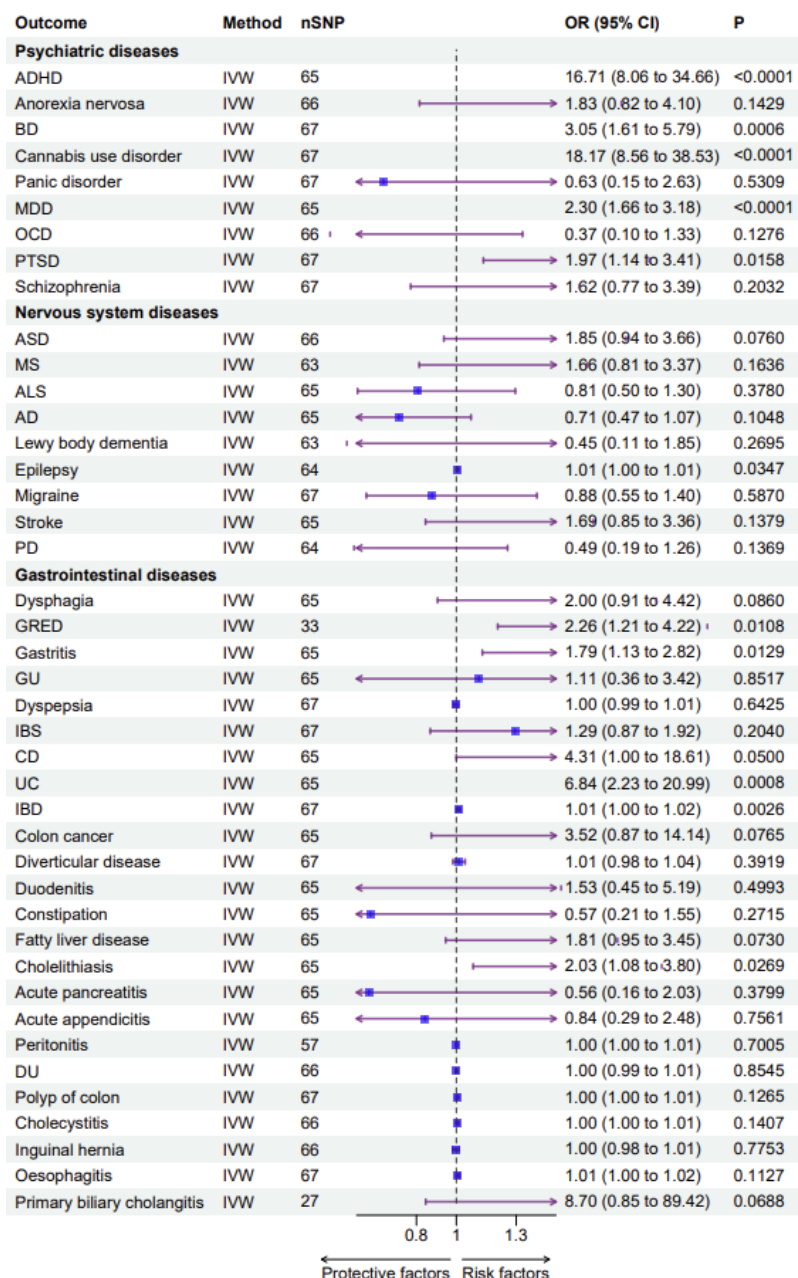
213

214 3.3 Causal inference analysis

215 Mendelian randomisation analysis suggested potential causal relationships between smoking
 216 and several diseases. Overall, smoking demonstrated significant positive causal effects on 6
 217 diseases after adjusting for multiple testing using the Benjamini–Hochberg (BH) method for P

218 value adjustment (Figure 2; Additional file: Supplementary Table S6), including ulcerative
219 colitis (UC) (IVW: odds ratio [OR] = 6.84, 95% confidence interval [CI] 2.23-20.99, $P < 0.001$),
220 inflammatory bowel disease (IBD) (IVW: OR = 1.01, 95% CI 1.00-1.02, $P = 0.003$), gastritis
221 (IVW: OR = 1.79, 95% CI 1.13-2.82, $P = 0.013$), post-traumatic stress disorder (PTSD) (IVW:
222 OR = 1.97, 95% CI 1.14-3.41, $P = 0.016$), cholelithiasis (IVW: OR = 2.03, 95% CI 1.08-3.80,
223 $P = 0.027$) and epilepsy (IVW: OR = 1.01, 95% CI 1.00-1.01, $P = 0.035$). Although Crohn's
224 disease (CD) and primary biliary cholangitis did not reach the 0.05 significance threshold, they
225 also exhibited positive causal tendencies. It is important to note that while smoking showed
226 positive causal effects on attention-deficit/hyperactivity disorder (ADHD), cannabis use
227 disorder, major depressive disorder (MDD), bipolar disorder (BD), and GRED, further
228 verification is required for this causal relationship, as indicated by the heterogeneity test and
229 MR-PRESSO test P values being < 0.05 , suggesting the presence of pleiotropy and less reliable
230 MR results.

231



232

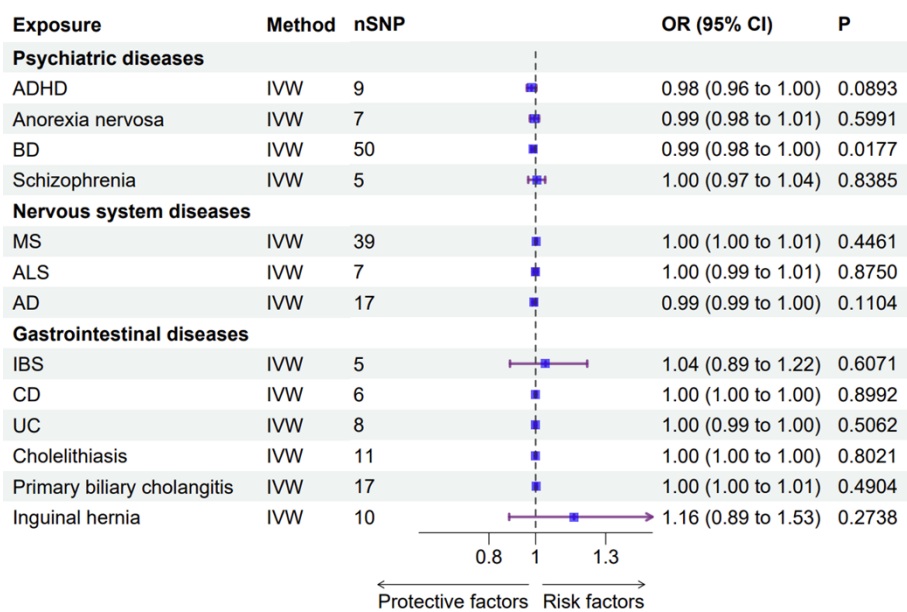
233 **Fig. 2 MR estimates for the effect of "ever smoked" on 42 diseases.** Estimates were derived by the
 234 inverse-variance weighted method. Reported P-values were adjusted for multiple testing using the Benjamini-
 235 Hochberg procedure. The solid lines represent 95% CI; CI, confidence interval; OR, odds ratio. ADHD, attention
 236 deficit hyperactivity disorder; BD, bipolar disorder; MDD, major depressive disorder; OCD, obsessive-compulsive
 237 disorder; PTSD, post-traumatic stress disorder; ASD, autism spectrum disorders; MS, multiple sclerosis; ALS,
 238 amyotrophic lateral sclerosis; AD, Alzheimer's disease; PD, Parkinson's disease; GRED, gastroesophageal reflux
 239 disease; GU, gastric ulcer; IBS, irritable bowel syndrome; CD, Crohn's disease; UC, ulcerative colitis; IBD,
 240 inflammatory bowel disease; DU, duodenal ulcer.

241

242 On the other hand, no disease demonstrated a significant causal effect on smoking after

243 adjusting for multiple testing (Figure 3; Additional file: Supplementary Table S7). Although
 244 BP (IVW: OR = 0.99, 95% CI 0.98-1.00, P = 0.018) presented a negative causal relationship
 245 with smoking, in the MR analysis, the result was less reliable due to the detection of horizontal
 246 pleiotropy. Leave-one-out analysis indicated that the causal relationships were not driven by
 247 any single SNP.

248



249

250 **Fig. 3 MR estimates for the effect of 42 diseases on "ever smoked"**. Estimates were derived by the
 251 inverse-variance weighted method. Reported P-values were adjusted for multiple testing using the Benjamini-
 252 Hochberg procedure. The solid lines represent 95% CI; CI, confidence interval; OR, odds ratio. ADHD, attention
 253 deficit hyperactivity disorder; BD, bipolar disorder; MS, multiple sclerosis; ALS, amyotrophic lateral sclerosis; AD,
 254 Alzheimer's disease; IBS, irritable bowel syndrome; CD, Crohn's disease; UC, ulcerative colitis.

255

256 Discussion

257 This investigation presents the most extensive compilation of publicly accessible GWAS data,
 258 enabling a comprehensive characterization of the genetic associations, pleiotropy and causality
 259 underlying smoking and neuropsychiatric and gastrointestinal diseases. Notably, our findings

260 unveil a significant genetic correlation between smoking and GRED. Moreover, our research
261 identifies six diseases in which smoking exhibits a causal relationship.

262 In this study, we report, for the first time, that among the 42 neuropsychiatric and
263 gastrointestinal diseases examined, only GERD exhibits a significant genetic association with
264 smoking. This finding aligns with previous epidemiological evidence suggesting a connection
265 between smoking and GRED (23–25). The genetic correlation between smoking and GRED
266 suggests the presence of shared genetic factors, indicating potential associations between the
267 two. However, further investigation is required to determine the precise mechanisms underlying
268 these relationships. Our pleiotropy analysis using MTAG revealed 358 SNPs jointly associated
269 with both smoking and GRED. However, we found no evidence supporting a significant causal
270 relationship between smoking and GRED in our MR analysis after horizontal pleiotropy and
271 heterogeneity tests. This study boasts a substantially larger sample size for GERD,
272 encompassing 129,080 cases and 473,524 controls, totaling 602,604 samples. This heightened
273 sample size enhances the statistical power of our results, rendering this study's conclusions
274 more reliable and compelling in refuting the previous finding(25) of a causal effect of smoking
275 on GERD (71,522 cases and 261,079 controls) and supporting the notion that the genetic
276 association is more likely attributable to pleiotropy.

277 Furthermore, we observed a potential genetic link between smoking and psychiatric disorders
278 such as schizophrenia, which corroborates the findings of previous studies(26,27). It is
279 noteworthy that both our study, conducted within a European population, and previous studies
280 conducted in East Asian populations did not find a significant causal relationship between
281 smoking and schizophrenia(28). Moreover, recent research suggests that the potential genetic

282 association between smoking and schizophrenia may stem from a local genetic correlation at
283 15q25(29). This discovery contributes to the mounting evidence supporting a genetic link
284 between smoking and schizophrenia arising from specific loci rather than causal effects.
285 Additionally, pathway clustering analysis revealed that the 513 pleiotropic genes shared
286 between smoking and the 42 diseases were enriched for pathways related to growth regulation
287 and synaptic function. This enrichment suggests that smoking may be implicated in disease
288 development by influencing neural and developmental processes. The enriched
289 neurodevelopmental pathways imply that smoking could disrupt crucial growth regulators
290 during neural maturation, thereby affecting disease risk. The synaptic pathways indicate that
291 smoking might influence neurotransmitter signaling and synaptic plasticity, thereby impacting
292 disease pathogenesis. By connecting smoking-associated genes to established disease pathways,
293 our integrative analysis yields new biological insights distinct from previous observational
294 studies. Nonetheless, experimental validations are necessary to confirm the functional roles of
295 the identified genes and pathways in mediating the relationships between smoking and diseases.
296 Our MR analysis provides compelling evidence of causal relationships between smoking and
297 several diseases, including UC, IBD, gastritis, PTSD, cholelithiasis, and epilepsy. These
298 findings underscore the unique strength of our study in comparison to previous research.
299 Consistent with prior studies, our MR analysis verifies the causal relationship between smoking
300 and cholelithiasis, IBD, and gastritis(30). Furthermore, our results align with a meta-analysis
301 reporting an increased risk of gallbladder disease associated with smoking(31). However, our
302 study diverges from previous research by furnishing robust genetic evidence that smoking does
303 not causally affect gastric and duodenal ulcers, irritable bowel syndrome, or CD. Our MR

304 analysis, with its larger sample size and rigorous statistical corrections, offers novel genetic
305 insights into the causal effects of smoking on diseases. In contrast, we did not detect significant
306 causal effects of the diseases on smoking after multiple testing correction. These null
307 associations indicate that while smoking may influence disease risk, the reverse direction of
308 disease liability contributing to smoking initiation is not supported by genetic evidence. Further
309 studies are warranted to replicate these findings and unravel the biological mechanisms
310 underlying the identified causal effects.

311 Several limitations merit attention in this study. First, the GWAS data primarily pertain to
312 European ancestry populations and may not be fully generalizable to other ethnic groups. Future
313 investigations involving more diverse populations are imperative to validate the findings.
314 Second, the causal inferences derived from MR analysis require further experimental validation
315 to elucidate the specific molecular mechanisms. In addition, the GWAS samples for smoking
316 and diseases may have partial overlap, which could introduce bias and lead to inflated causal
317 associations. Although we endeavored to maximize the independence between exposure and
318 outcome samples during MR analysis, it is crucial to consider the potential sample overlap
319 when interpreting the causal effects. Additionally, the study may be subject to limitations
320 inherent to the statistical methodology, making it challenging to determine weaker causal
321 relationships between smoking and certain diseases. Our findings call for further research into
322 biological pathways to comprehensively delineate the causality between smoking and various
323 diseases, ultimately advancing precision medicine. Despite these limitations, our study offers
324 comprehensive insights into the potential causal effects of smoking on a wide range of disease
325 endpoints within both the neuropsychiatric and gastrointestinal systems. Further rigorous

326 investigations in the future can aid in validating and expanding upon the findings presented
327 here.

328 **Conclusions**

329 Overall, by integrating multiple genetic approaches, this study unveils the shared genetics and
330 potential pathogenic effects linking smoking to neuropsychiatric and gastrointestinal diseases,
331 providing crucial insights for subsequent research in this domain. Moreover, these findings hold
332 promise for advancing clinical translational applications that aim to elucidate the intricate
333 relationships between smoking and various diseases.

334 **List of Abbreviations**

335 **MR:** Mendelian randomisation
336 **GWAS:** Genome-wide association study
337 **SNP:** Single-nucleotide polymorphisms
338 **GRED:** Gastroesophageal reflux disease
339 **PGC:** The Psychiatric Genomics Consortium
340 **LDSC:** The linkage disequilibrium score regression
341 **MTAG:** Multitrait analysis of GWAS
342 **FDR:** False discovery rate
343 **GO:** Gene Ontology
344 **IVW:** Inverse variance weighted
345 **BH:** Benjamini–Hochberg
346 **OR:** Odds ratio

347 **CI:** Confidence interval

348 **UC:** Ulcerative colitis

349 **IBD:** Inflammatory bowel disease

350 **PTSD:** Post-traumatic stress disorder

351 **CD:** Crohn's disease

352 **ADHD:** Attention-deficit/hyperactivity disorder

353 **MDD:** Major depressive disorder

354 **BD:** Bipolar disorder

355

356 **Declarations**

357 **Ethics approval and consent to participate**

358 Not applicable.

359 **Consent for publication**

360 Not applicable.

361 **Availability of data and materials**

362 The GWAS datasets supporting the conclusions of this article are available in the GWAS

363 Catalog (<https://www.ebi.ac.uk/gwas/>), the Psychiatric Genomics Consortium

364 (<https://www.med.unc.edu/pgc/>) and the MRC-IEU OpenGWAS database

365 (<https://gwas.mrcieu.ac.uk/>). The Migraine GWAS dataset provided by Hautakangas et al.(32)

366 can be obtained by contacting the International Headache Genetics Consortium. Data analysis

367 was performed using R version 4.2.2. LDSC (<https://github.com/bulik/ldsc/>), MTAG
368 (<https://github.com/JonJala/mtag/>) and MAGMA(<http://ctglab.nl/software/magma>) were used
369 for analysis, and Metascape (<https://metascape.org/>) was used for gene ontology enrichment
370 analysis. This study was not preregistered.

371 **Competing interests**

372 The authors declare that they have no competing interests.

373 **Funding**

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

378 **Authors' Information**

379 Affiliations

380 Bioinformatics Centre, National Clinical Research Centre for Geriatric Disorders, Department
381 of Geriatrics, Xiangya Hospital, Central South University, Changsha, Hunan 410008, China.

382 Jiayi Xiong, Zheng Wang, Yuanfeng Huang, Shiyu Zhang, Jiaqi Yang, Shuo Gao, Tianyang
383 Wang, Jinchun Li, Guihu Zhao & Bin Li

384 Centre for Medical Genetics & Hunan Key Laboratory, School of Life Sciences, Central
385 South University, Changsha, Hunan 410008, China.

386 Jinchun Li

387 Department of Neurology, Xiangya Hospital, Central South University, Changsha, Hunan

388 410008, China.

389 Zheng Wang, Jinchen Li, Guihu Zhao & Bin Li

390 Hunan Key Laboratory of Molecular Precision Medicine, Xiangya Hospital, Central South

391 University, Changsha, Hunan 410008, China.

392 Zheng Wang

393 Xiangya School of Medicine, Central South University, Changsha, Hunan 410013, China.

394 Shiyu Zhang & Tianyang Wang

395 Bioinformatics Centre, Furong Laboratory, Changsha, Hunan 410008, China.

396 Jinchen Li

397 Department of Orthopedics, Xiangya Hospital, Central South University, Changsha 410008,

398 Hunan, China.

399 Guang Yang

400 **Contributions**

401 JCL, GHZ and BL designed and supervised the study, while JYX performed the majority of the

402 analysis and drafted the manuscript. BL provided mentorship, and all authors contributed to the

403 study, interpreted the findings, and reviewed the manuscript. The final version of the

404 manuscript was approved by all authors before publication.

405 **Corresponding author**

406 Correspondence to: Jinchen Li, Guihu Zhao, Bin Li

407 Address: No.87 Xiangya Road, Changsha, Hunan, China

408 Telephone: 0731-89752068

409 E-mail address: lijinchen@csu.edu.cn, ghzhao@csu.edu.cn, lebin001@csu.edu.cn

410 Acknowledgements

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

415 References

- 416 1. Cornelius ME, Loretan CG, Jamal A, Lynn BCD, Mayer M, Alcantara IC, et al. Tobacco
417 Product Use Among Adults – United States, 202. 2023;72(18).
- 418 2. Kamceva G, Arsova-Sarafinovska Z, Ruskovska T, Zdravkovska M, Kamceva-Panova L,
419 Stikova E. Cigarette Smoking and Oxidative Stress in Patients with Coronary Artery Disease.
420 Open Access Maced J Med Sci. 2016 Oct 28;4(4):636–40.
- 421 3. Barkhuizen W, Dudbridge F, Ronald A. Genetic overlap and causal associations between
422 smoking behaviours and mental health. Sci Rep. 2021 Jul 21;11(1):14871.
- 423 4. Papoutsopoulou S, Satsangi J, Campbell BJ, Probert CS. Review article: impact of cigarette
424 smoking on intestinal inflammation—direct and indirect mechanisms. Alimentary
425 Pharmacology & Therapeutics. 2020;51(12):1268–85.
- 426 5. Winsvold BS, Harder AVE, Ran C, Chalmer MA, Dalmaso MC, Ferkingstad E, et al. Cluster
427 Headache Genomewide Association Study and Meta-Analysis Identifies Eight Loci and
428 Implicates Smoking as Causal Risk Factor. Ann Neurol. 2023 Jul 24;10.1002/ana.26743.
- 429 6. Vandebergh M, Goris A. Smoking and multiple sclerosis risk: a Mendelian randomization

- 430 study. *J Neurol*. 2020 Oct;267(10):3083–91.
- 431 7. Larsson SC, Carter P, Kar S, Vithayathil M, Mason AM, Michaëlsson K, et al. Smoking,
432 alcohol consumption, and cancer: A mendelian randomisation study in UK Biobank and
433 international genetic consortia participants. *PLoS Med*. 2020 Jul 23;17(7):e1003178.
- 434 8. Larsson SC, Mason AM, Bäck M, Klarin D, Damrauer SM, Million Veteran Program, et al.
435 Genetic predisposition to smoking in relation to 14 cardiovascular diseases. *European Heart*
436 *Journal*. 2020 Sep 14;41(35):3304–10.
- 437 9. Larsson SC, Burgess S. Appraising the causal role of smoking in multiple diseases: A
438 systematic review and meta-analysis of Mendelian randomization studies. *eBioMedicine*. 2022
439 Aug;82:104154.
- 440 10. Cryan JF, O’Riordan KJ, Cowan CSM, Sandhu KV, Bastiaanssen TFS, Boehme M, et al. The
441 Microbiota-Gut-Brain Axis. *Physiological Reviews*. 2019 Oct 1;99(4):1877–2013.
- 442 11. Gong W, Guo P, Li Y, Liu L, Yan R, Liu S, et al. Role of the Gut-Brain Axis in the Shared
443 Genetic Etiology Between Gastrointestinal Tract Diseases and Psychiatric Disorders: A
444 Genome-Wide Pleiotropic Analysis. *JAMA Psychiatry*. 2023 Apr 1;80(4):360.
- 445 12. Margolis KG, Cryan JF, Mayer EA. The Microbiota-Gut-Brain Axis: From Motility to Mood.
446 *Gastroenterology*. 2021 Apr;160(5):1486–501.
- 447 13. Mappin-Kasirer B, Pan H, Lewington S, Kizza J, Gray R, Clarke R, et al. Tobacco smoking
448 and the risk of Parkinson disease: A 65-year follow-up of 30,000 male British doctors.

- 449 Neurology. 2020 May 19;94(20):e2132–8.
- 450 14. Botteri E, Borroni E, Sloan EK, Bagnardi V, Bosetti C, Peveri G, et al. Smoking and Colorectal
451 Cancer Risk, Overall and by Molecular Subtypes: A Meta-Analysis. *Am J Gastroenterol*. 2020
452 Dec;115(12):1940–9.
- 453 15. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Bulik-Sullivan BK,
454 Loh PR, Finucane HK, Ripke S, Yang J, et al. LD Score regression distinguishes confounding
455 from polygenicity in genome-wide association studies. *Nat Genet*. 2015 Mar;47(3):291–5.
- 456 16. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis
457 of genome-wide association summary statistics using MTAG. *Nat Genet*. 2018 Feb;50(2):229–
458 37.
- 459 17. Zeng R, Jiang R, Huang W, Wang J, Zhang L, Ma Y, et al. Dissecting shared genetic
460 architecture between obesity and multiple sclerosis. *eBioMedicine*. 2023 Jul;93:104647.
- 461 18. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of
462 GWAS data. *PLoS Comput Biol*. 2015 Apr 17;11(4):e1004219.
- 463 19. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape
464 provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*.
465 2019 Apr 3;10(1):1523.
- 466 20. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R Package for Comparing Biological
467 Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*. 2012

- 468 May;16(5):284–7.
- 469 21. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple
470 Genetic Variants Using Summarized Data. *Genet Epidemiol.* 2013 Nov;37(7):658–65.
- 471 22. Lee CH, Cook S, Lee JS, Han B. Comparison of Two Meta-Analysis Methods: Inverse-
472 Variance-Weighted Average and Weighted Sum of Z-Scores. *Genomics Inform.*
473 2016;14(4):173.
- 474 23. Sharma P. Barrett Esophagus: A Review. *JAMA.* 2022 Aug 16;328(7):663.
- 475 24. Maret-Ouda J, Markar SR, Lagergren J. Gastroesophageal Reflux Disease: A Review. *JAMA.*
476 2020 Dec 22;324(24):2536.
- 477 25. Yuan S, Larsson SC. Adiposity, diabetes, lifestyle factors and risk of gastroesophageal reflux
478 disease: a Mendelian randomization study. *Eur J Epidemiol.* 2022 Jul;37(7):747–54.
- 479 26. Hahad O, Daiber A, Michal M, Kuntic M, Lieb K, Beutel M, et al. Smoking and
480 Neuropsychiatric Disease—Associations and Underlying Mechanisms. *IJMS.* 2021 Jul
481 6;22(14):7272.
- 482 27. Hartz SM, Horton AC, Hancock DB, Baker TB, Caporaso NE, Chen LS, et al. Genetic
483 correlation between smoking behaviors and schizophrenia. *Schizophrenia Research.* 2018
484 Apr;194:86–90.
- 485 28. Chen J, Chen R, Xiang S, Li N, Gao C, Wu C, et al. Cigarette smoking and schizophrenia:
486 Mendelian randomisation study. *Br J Psychiatry.* 2021 Feb;218(2):98–103.

- 487 29. Gerring ZF, Thorp JG, Gamazon ER, Derks EM. A Local Genetic Correlation Analysis
488 Provides Biological Insights Into the Shared Genetic Architecture of Psychiatric and Substance
489 Use Phenotypes. *Biological Psychiatry*. 2022 Oct;92(7):583–91.
- 490 30. Yuan S, Chen J, Ruan X, Sun Y, Zhang K, Wang X, et al. Smoking, alcohol consumption, and
491 24 gastrointestinal diseases: Mendelian randomization analysis. *eLife*. 12:e84051.
- 492 31. Aune D, Vatten LJ, Boffetta P. Tobacco smoking and the risk of gallbladder disease. *Eur J*
493 *Epidemiol*. 2016 Jul;31(7):643–53.
- 494 32. Hautakangas H, Winsvold BS, Ruotsalainen SE, Bjornsdottir G, Harder AVE, Kogelman LJA,
495 et al. Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-
496 specific risk alleles. *Nat Genet*. 2022 Feb;54(2):152–60.

497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514

