

1 **Phenome-wide association study to explore the long-term symptoms after**  
2 **infection with novel coronavirus in the UK Biobank**

3 **Running title:** Phenome-wide association study for infection with novel coronavirus

4 Kai Zhang<sup>1,3</sup>, Xiaowen Liu<sup>1,3</sup>, Ping Fu<sup>1,3</sup>, Yingqi Zhao<sup>1,3</sup>, Qingqing Yu<sup>1,3</sup>, Shuling Liu<sup>1,3</sup>, Fuzhong  
5 Xue<sup>1,2,3\*</sup> (0000-0003-0378-7956)

6 1. Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong  
7 University, Jinan, 250012, P.R. China.

8 2. Institute for Medical Dataology, Shandong University, Jinan, 250002, P.R. China.

9 3. National Institute of Health Data Science of China, Jinan, 250002, P.R. China.

10 **\*Correspondence to:** Fuzhong Xue, Ph.D., Department of Biostatistics, School of Public Health,  
11 Cheeloo College of Medicine, Shandong University, No.44 Wenhuxi Road, Jinan, 250012, China.  
12 (Email: xuefzh@sdu.edu.cn; Tel: (+86)-531-88380280; Fax: (+86)-531-88380280).

13 **Competing interests:** The authors declare no potential conflicts of interest.

## 14 **Abstract**

### 15 **Background**

16 Observational research studies have shown that even after the acute phase, severe acute respiratory  
17 syndrome coronavirus 2 (SARS-CoV-2) can affect patients, and increase the risk of cardiovascular,  
18 mental, metabolic, and other disorders. However, the spectrum of diseases for individuals with a  
19 genetic predisposition to COVID-19 remains unclear.

### 20 **Methods**

21 We leveraged individual-level data from UK Biobank to implement a phenome-wide association  
22 study to explore the relationships between COVID-19 and 1061 diseases. Then, the inverse-variance  
23 weighted (IVW) method was adopted with summary-level data from global consortiums as sensitivity  
24 analyses combined with other MR methods with different model assumptions to identify robust  
25 associations.

### 26 **Findings**

27 The PheWAS found severe respiratory, hospitalized, and susceptibility COVID-19 had detrimental  
28 effects on 36, 37, and 51 kinds of diseases, separately. The IVW test found severe respiratory COVID-  
29 19 had detrimental effects on breast cancer [OR 95% CI: 1.065 (1.000-1.133)], pan-cancer [OR 95%  
30 CI: 1.002 (1.000-1.004)], and Alzheimer's disease [OR 95% CI: 1.042 (1.005-1.081)], etc.  
31 Hospitalized COVID-19 had detrimental effects on ischemic stroke (IS) [OR 95%CI: 1.049 (1.001-  
32 1.100)], breast cancer [OR 95%CI: 1.139 (1.011-1.283)], and pan-cancer [OR 95%CI: 1.003 (1.000-  
33 1.006)], etc. Susceptibility COVID-19 had detrimental effects on deep vein thrombosis (DVT) of  
34 lower extremities [OR 95%CI: 2.392 (1.167-4.902)], venous thromboembolism [OR 95%CI: 1.962  
35 (1.115-3.453)], pulmonary heart disease/diseases of pulmonary circulation [OR 95%CI: 1.767  
36 (1.142-2.733)], IS (large artery atherosclerosis) [OR 95%CI: 1.405 (1.025-1.927)], myocardial  
37 infarction [OR 95%CI: 1.235 (1.012-1.509)], heart failure [OR 95%CI: 1.140 (1.009-1.287)], etc.

### 38 **Interpretation**

39 This study describes the extensive link between genetically determined COVID-19 and a broad range

40 of diseases, especially those of the circulatory system, neuropsychiatric system, neoplasms, immune  
41 system, and digestive systems. Early detection and management of post-COVID-19 conditions could  
42 be tremendously beneficial to public health.

43 **Funding**

44 This work was supported by the National Natural Science Foundation of China (81773547 and  
45 82173625) and the National Key Research and Development Program (2020YFC2003500).

46 **Keywords** post-COVID-19 condition, phenome-wide association study (PheWAS), Mendelian  
47 randomization, causality

## 48 **1. Introduction**

49 The coronavirus disease 2019 (COVID-19) has spread across the world. As of 31 December 2022,  
50 more than 657 million confirmed cases of COVID-19 have been recorded worldwide, and more than  
51 6.68 million deaths have been reported by the World Health Organization (WHO). Much research  
52 shows that even after the acute phase, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-  
53 2) can affect patients and impair their quality of life. Recently, International organizations and  
54 Institutes have drawn attention to an increasing number of people experiencing health consequences  
55 following the acute phase of SARS-CoV-2 infection and are calling for research into the risk factors,  
56 clinical features, diagnosis, management, and outcomes [1–5]. Indeed, even the terminology of the  
57 condition is debated with variable terms and definitions for the post-COVID-19 condition including  
58 long COVID, long-haul COVID, long COVID-19 condition, post-acute sequelae of SARS-CoV-2  
59 infection (PASC), or post-COVID-19 condition (the term used by WHO) [6–8].

60 Long-term health consequences of COVID-19 remain unknown, but most study report that the  
61 spectrum of long-lasting symptoms is wide and varies from mild discomfort to severe adverse effects  
62 on physical, cognitive, and psychosocial health, with important wider implications on functioning,  
63 including employment and school attendance [9]. Multiple studies from different countries found that  
64 many individuals experienced persistent symptoms 6 months after COVID-19, with fatigue or muscle  
65 weakness, sleep difficulties, and anxiety or depression among the most common sequelae [10–12]. A  
66 recent study suggests that although most COVID-19 survivors recover both physically and  
67 functionally a year after acute infection, some still experience problems with mobility, pain or  
68 discomfort, and anxiety or depression compared with non-COVID-19 controls [13]. The data  
69 emerging from the controlled studies are in agreement with the earlier reports. A recent analysis of  
70 the data from over 87 million electronic health records demonstrated that more than one in three  
71 individuals had one or more features of post-COVID-19 condition recorded between 3 and 6 months  
72 after a diagnosis of COVID-19, which was significantly higher when compared with individuals with  
73 influenza [14–18]. Disease severity, female sex, and age were associated with a higher risk of post-  
74 COVID-19 condition development.

75 However, observational evidence is less likely to account for unknown confounding factors and

76 reverse causation bias [19]. It is also important to note that most data regarding post-COVID-19  
77 conditions have been generated before the condition definition announcement. Thus, earlier studies  
78 may not fit the proposed definition criteria. In addition, these studies have often focused on a single  
79 type of disease. The wide range of diseases associated with COVID-19 requires an urgent systematic  
80 assessment.

81 A phenome-wide association study (PheWAS) [20] combined with a two-sample Mendelian  
82 randomization (MR) approach was suitable for use in examining genetic evidence between the  
83 COVID-19 phenotypes and diseases, providing a more comprehensive assessment of post-COVID-  
84 19 condition associated human complex diseases. The PheWAS paradigm was introduced in 2010 as  
85 an approach that scans across a range of phenotypes, similar to what was accomplished in genome-  
86 wide association studies (GWAS) [20]. Studies that adopted PheWAS analysis have been used to  
87 understand the beneficial or detrimental effects of body mass index (BMI), serum urate, age at  
88 menarche, smoking intensity, ABO blood groups, and C-reactive protein (CRP) on human complex  
89 phenotypes [21–26].

90 This study leveraged both PheWAS approaches and MR to identify the spectrum of human  
91 diseases associated with genetically predisposed COVID-19. In addition, several complementary  
92 two-sample MR approaches were used to identify robust associations.

## 93 **2. Materials and methods**

### 94 ***2.1 Study Population***

95 UK Biobank contains in-depth genetic and health information on over 500,000 participants.[27] This  
96 prospective study collected data including blood, urine, and saliva laboratory tests and questionnaires,  
97 physical measurements, and genome-wide genotyping information.[27] This information is linked to  
98 national primary care and inpatient records, cancer diagnosis, and death registration data for  
99 longitudinal follow-up.[27] Participants who withdrew from the study or were lost to follow-up, those  
100 lacking genotype information, self-reported sex data that did not match genetic records, sex  
101 chromosome aneuploidy, >10 putative third-degree relatives in the kinship table, excessive  
102 heterozygosity (top 1%), non-Caucasian ancestry, and missing key covariates (age, sex, BMI, smoke,

103 and drink status) were excluded from the analysis. Genetic quality control was performed centrally  
104 by the UK Biobank.[28] Detailed information about this cohort can be found on the Biobank website  
105 (<https://www.ukbiobank.ac.uk/>). For PheWAS, 455,509 participants with genotypes in the UK  
106 Biobank were included.

## 107 **2.2 GWAS summary-level data for COVID-19**

108 We evaluated three COVID-19 phenotypes: severe respiratory, hospitalized, and susceptibility to  
109 COVID-19. These data were derived from Release 7 of the GWAS meta-analysis conducted by the  
110 COVID-19 Host Genetics Initiative (HGI). (<https://www.covid19hg.org/results/r7/>) This is an  
111 international genetics collaboration that aims to uncover the genetic determinants of COVID-19  
112 susceptibility, severity, and outcomes.[29] For the evaluated COVID-19 outcomes, controls were  
113 genetically ancestry-matched individuals without SARS-CoV-2 infection.[29] In our analyses, we  
114 used GWAS summary statistics from the comparison between cases and control groups of each  
115 exposure. The severe respiratory COVID-19 outcome resulted from the comparison between patients  
116 with very severe respiratory failure secondary to COVID-19 ( $n = 18,152$ ) vs controls ( $n = 1,145,546$ ).  
117 The hospitalized COVID-19 data were generated from the comparison of patients with a laboratory-  
118 confirmed SARS-CoV-2 infection that was hospitalized due to COVID-19 symptoms ( $n = 44,986$ ) vs  
119 controls ( $n = 2,356,386$ ). Finally, the susceptibility COVID-19 analysis was conducted by comparing  
120 159,840 individuals reporting SARS-CoV-2 infection with 2,782,977 controls. Information regarding  
121 SARS-CoV-2 infection was derived from a laboratory test, electronic health record, clinically  
122 confirmed COVID-19, and self-reported COVID-19 (e.g. by questionnaire).[29] Single nucleotide  
123 polymorphisms (SNPs) were filtered according to the following criteria: 1) a genome-wide threshold  
124 of significance ( $P < 5 \times 10^{-8}$ ) and 2) linkage disequilibrium (LD) defined as  $r^2 < 0.01$  and clump  
125 window  $> 10,000$  kb.[30] (**Supplementary Table S1**)

## 126 **2.3 Phenome Construction**

127 Phenotypes were constructed using the PheCODE system, which was developed to combine one or  
128 more relevant International Classification of Diseases (ICD) codes into different disease groups,  
129 allowing unbiased interrogation of multiple phenotypes in an EMR-based cohort.[31] To construct  
130 the phenotypes, a map designed for large biobanks was used to match ICD-9/10 codes to the

131 “phecode.”[32] The PheCODE system also provided a scheme to automatically exclude patients with  
132 similar or related diseases from the controls.[22] Primary and secondary ICD-9/10 codes in the UK  
133 Biobank were pooled from hospital records, cancer registry, and death registry data, and mapped into  
134 phecodes using the R package, “PheWAS”.[33] The phecode-mapping rules, as well as the excluding  
135 standards for each code, are available at [https://phewascatalog.org/phecodes\\_icd10](https://phewascatalog.org/phecodes_icd10). Phecodes with  
136 more than 200 cases were included in the final analysis. A total of 1061 phenotypes were utilized in  
137 the PheWAS and divided into 17 categories: circulatory, congenital, dermatologic, digestive,  
138 endocrine/metabolic, genitourinary, hematopoietic, infectious disease, injuries and poisonings,  
139 mental disorders, musculoskeletal, neoplasms, neurological, pregnancy complications, respiratory,  
140 sense organ, and symptom disorders. (**Supplementary Table S2**)

## 141 ***2.4 Statistical Analyses***

### 142 ***2.4.1 Phenome-wide association study***

143 The PRS was determined using the equation below.

$$144 \beta_i = \ln(OR_i) \quad (1)$$

$$145 PRS = \beta_1 \times SNP_1 + \beta_2 \times SNP_2 + \dots + \beta_i \times SNP_i \dots + \beta_n \times SNP_n \quad (2)$$

146 This PRS calculation method assumes an additive genotype model that an individual may have  
147 0, 1, or 2 risk-increasing alleles. The sum of the risk alleles is calculated using the effect size from  
148 the variant-exposure association, defined as  $\beta$ , which is obtained from the discovery of GWAS [34],  
149 and  $n$  is defined as the total number of SNP included in the model. Z-transformation was performed  
150 on GRS to ensure that the final result corresponded to the disease risk of 1-SD PRS increments. All  
151 models assumed a multivariable logistic model and adjusted for age, sex, BMI, smoking status, and  
152 drinking status. The “PheWAS” package was applied to test associations between genetically  
153 determined COVID-19 and hundreds of phenotypes in the UK Biobank cohort. The signals that  
154 passed the Bonferroni correction in the PheWAS analysis were used for further two-sample MR  
155 analyses.

### 156 ***2.4.2 Two-Sample Mendelian Randomization***

157 For the candidate phenotypes detected using PheWAS, two-sample MR analyses were performed in

158 parallel to assess the robustness of causal findings. Summary-level datasets for these diseases were  
159 acquired from global consortiums and excluded the UK Biobank-related data sources to avoid  
160 potential bias from sample overlap. A total of 207 disease phenotypes in the finding phase have  
161 available GWAS datasets. These summary-level statistics was shown in **Supplementary Table S3**.  
162 The  $F$ -statistic of  $>10$  could be considered a strong IV.

163 The IVW approach was used as the primary analysis with the assumption of no invalid IVs.[35]  
164 When the number of IVs was less than or equal to three, we applied the fixed-effects model to estimate  
165 the causal effect between exposure and outcome, otherwise, the random-effects model was used.  
166 Furthermore, the simple median (SME) and weighted-median estimator (WME) allow more powerful  
167 genetic variants to contribute more. It could be obtained by weighting the contribution of each genetic  
168 variant according to the inverse variance associated with the result. Even if up to 50% of the IVs are  
169 invalid, the estimator is consistent. [36] MR-Egger evaluates whether the pleiotropic effect of genetic  
170 variants on the result is different from zero on average [37]. MR-Egger is similar to IVW, but the  
171 former adjusts IVW analysis by allowing non-zero intercepts, namely, allowing horizontal pleiotropic  
172 effects. Even if all of the genetic variants violate IV assumptions 2, MR-Egger also returns an  
173 unbiased estimate of causal effects. [37] In MR-Egger regression, the estimate of intercept can be  
174 interpreted as an estimate of the average pleiotropy of all genetic variants, and the slope coefficient  
175 provides an estimate of the bias of the causal effect. [37] For binary variables, the MR estimates were  
176 reported as odds ratios (ORs) which can be interpreted as the risk increase of outcome per unit  
177 increase in log odds of exposure.

178 Phenotypes that passed the sensitivity analysis were defined as robust associations. All analyses  
179 were two-tailed and performed using R software (Version 3.6.3) with the ‘phewas’ and  
180 ‘TwoSampleMR’ packages.

### 181 **3. Results**

182 **Figure 1A and Figure 1B** showed the results of the PheWAS analysis between severe respiratory  
183 COVID-19 and 1061 binary phenotypes in the UK Biobank. **Figure 1A** showed that severe  
184 respiratory COVID-19 had detrimental effects on 36 diseases. These diseases were categorized as  
185 neoplasms (n=6), congenital anomalies (n=5), circulatory system (n=3), dermatologic (n=3),



186 digestive (n=3), infectious diseases (n=3), endocrine/metabolic (n=2), injuries & poisonings (n=2),  
187 musculoskeletal (n=2), and one kind of mental disorders, neurological, respiratory, sense organs,  
188 symptoms, respectively. No associations were evident between severe respiratory COVID-19 and  
189 hematopoietic disorders and pregnancy complications.

190 The odds ratios are shown in **Figure 2**. For circulatory system diseases, the odds ratio for the  
191 association between severe respiratory COVID-19 and atherosclerosis of aorta, other specified  
192 cardiac dysrhythmias, varicose veins of lower extremity, varicose veins were 1.160 (95% Confidence  
193 interval, 95% CI,1.028–1.293), 1.033 (95% CI,1.001–1.065), 1.022 (95% CI,1.004–1.040), and  
194 1.020 (95% CI,1.002–1.037), respectively. For congenital anomalies, the odds ratios for the  
195 association between severe respiratory COVID-19 and congenital anomalies of intestine, lower  
196 gastrointestinal congenital anomalies, congenital deformities of feet, digestive congenital anomalies,  
197 upper gastrointestinal congenital anomalies were 1.199 (95% CI, 1.082–1.317), 1.151 (95% CI,  
198 1.057–1.245), 1.137 (95% CI, 1.016–1.259), 1.127 (95% CI, 1.061–1.194), and 1.101 (95% CI,  
199 1.007–1.195), respectively. For neoplasms, the odds ratios for the association between severe  
200 respiratory COVID-19 and other benign neoplasm of connective and other soft tissue, cancer of  
201 esophagus, benign neoplasm of brain and other parts of nervous system, cancer of stomach, benign  
202 neoplasm of brain/cranial nerves/meninges, breast cancer, encephalitis were 1.124 (95% CI,  
203 1.048–1.199), 1.092 (95% CI, 1.031–1.153), 1.078 (95% CI, 1.020–1.135), 1.075 (95% CI,  
204 1.007–1.144), 1.066 (95% CI, 1.006–1.126), 1.021 (95% CI, 1.001–1.041), and 1.104 (95% CI,  
205 1.006–1.202), respectively.

206 **Figure 1C** showed that hospitalized COVID-19 had detrimental effects on 37 diseases. These  
207 diseases were categorized as neoplasms (n=7), congenital anomalies (n=4), genitourinary(n=4),  
208 circulatory system (n=3), dermatologic (n=3), digestive (n=3), endocrine/metabolic (n=3), injuries &  
209 poisonings (n=2), musculoskeletal (n=2), neurological (n=2), mental disorders (n=1), respiratory  
210 (n=1), and sense organs (n=1). The odds ratios are shown in **Figure 3**. For circulatory system diseases,  
211 the odds ratio for the association between hospitalized COVID-19 and cancer of esophagus, other  
212 benign neoplasm of connective and other soft tissue, breast cancer, breast cancer [female], lipoma of  
213 skin and subcutaneous tissue, cancer of prostate, malignant neoplasm of female breast were 1.104  
214 (95% CI, 1.043–1.165), 1.087 (95% CI, 1.010–1.163), 1.032 (95% CI, 1.012–1.051), 1.030 (95% CI,

215 1.010–1.049), 1.029 (95% CI, 1.001–1.057), 1.025 (95% CI, 1.005–1.046), and 1.023 (95% CI,  
216 1.001–1.044), respectively. For circulatory system, the odds ratio for the association between  
217 hospitalized COVID-19 and atherosclerosis of aorta, varicose veins of lower extremity, varicose veins  
218 were 1.197 (95% CI, 1.064–1.329), 1.021 (95% CI, 1.004–1.039), and 1.019 (95% CI, 1.001–1.036),  
219 respectively.

220 **Figure 2E** showed that susceptibility to COVID-19 had detrimental effects on 51 diseases. These  
221 diseases were categorized as endocrine/metabolic (n=10), circulatory system (n=9), neoplasms (n=7),  
222 dermatologic (n=6), digestive (n=4), musculoskeletal (n=3), respiratory (n=3), genitourinary (n=2),  
223 sense organs (n=2), symptoms (n=2), injuries & poisonings (n=1), neurological (n=1), and congenital  
224 anomalies (n=1). The odds ratios are shown in **Figure 4**. For endocrine/metabolic system diseases,  
225 the odds ratio for the association between COVID-19 infection and mineral deficiency, crystal  
226 arthropathies, type 2 diabetes with ophthalmic manifestations, vitamin D deficiency, vitamin  
227 deficiency, vitamin B-complex deficiencies, electrolyte imbalance, disorders of fluid/electrolyte/ and  
228 acid–base balance, diabetes mellitus, type 2 diabetes were 1.118 (95% CI, 1.016–1.221), 1.097 (95%  
229 CI, 1.007–1.186), 1.096 (95% CI, 1.041–1.151), 1.078 (95% CI, 1.013–1.143), 1.067 (95% CI,  
230 1.023–1.111), 1.063 (95% CI, 1.003–1.122), 1.034 (95% CI, 1.009–1.060), 1.025 (95% CI,  
231 1.005–1.046), 1.020 (95% CI, 1.006–1.034), and 1.019 (95% CI, 1.005–1.033), respectively. For  
232 circulatory system, the odds ratio for the association between COVID-19 infection and atherosclerosis  
233 of aorta, other venous embolism and thrombosis, phlebitis and thrombophlebitis of lower extremities,  
234 phlebitis and thrombophlebitis, cardiac arrest and ventricular fibrillation, circulatory disease, other  
235 disorders of circulatory system, precordial pain, cardiac dysrhythmias, other specified congenital  
236 anomalies of kidney were 1.164 (95% CI, 1.031–1.296), 1.153 (95% CI, 1.034–1.271), 1.117 (95%  
237 CI, 1.057–1.177), 1.085 (95% CI, 1.034–1.136), 1.068 (95% CI, 1.007–1.129), 1.034 (95% CI,  
238 1.019–1.048), 1.032 (95% CI, 1.018–1.046), 1.032 (95% CI, 1.002–1.061), 1.016 (95% CI,  
239 1.004–1.028), and 1.121 (95% CI, 1.025–1.216), respectively.

240 As shown in **Figure 5**, IVW test found a statistically significant association between critical  
241 COVID-19 and 26 diseases, including circulatory system (n=3): ischemic stroke (small–vessel) (OR,  
242 95% CI: 0.941, 0.887–0.998), conduction disorders (OR, 95% CI: 0.935, 0.880 –0.992),  
243 atrioventricular (AV)-block (OR, 95% CI: 0.900, 0.830–0.976); digestive system (n=5):

244 gastroesophageal reflux disease (OR, 95% CI: 1.022, 1.004-1.040), hernia (OR, 95% CI: 0.942,  
245 0.914-0.970), inguinal hernia (OR, 95% CI: 0.932, 0.898-0.968), ulcerative colitis (OR, 95% CI:  
246 0.924, 0.860-0.992), and Crohn's disease (OR, 95% CI: 0.327, 0.211-0.505); neoplasms (n=7):  
247 malignant neoplasm of breast (HER-positive) (OR, 95% CI: 1.065, 1.000-1.133), malignant neoplasm  
248 of breast (HER2-negative, all cancers excluded) (OR, 95% CI: 1.064, 1.012-1.118), malignant  
249 neoplasm of breast (HER2-negative) (OR, 95% CI: 1.061, 1.011-1.114), malignant neoplasm of  
250 breast (all cancers excluded) (OR, 95% CI: 1.057, 1.007- 1.110), pan cancer (OR, 95% CI: 1.002,  
251 1.000-1.004), ovarian cancer (OR, 95% CI: 1.001, 1.000-1.002), oesophageal cancer (OR, 95% CI:  
252 1.000, 1.000 -1.001); neuropsychiatric system (n=8): Alzheimer's disease or family history of  
253 Alzheimer's disease (OR, 95% CI: 1.042, 1.005-1.081), father's history of Alzheimer's disease (OR,  
254 95% CI: 1.041, 1.005 -1.077), paternal history of Alzheimer's disease (OR, 95% CI: 1.003, 1.000-  
255 1.006), family history of Alzheimer's disease (OR, 95% CI: 0.923, 0.860-0.990), major depressive  
256 disorder (OR, 95% CI: 0.839, 0.730-0.965), hyperkinetic disorders (more controls excluded) (OR,  
257 95% CI: 0.744, 0.615-0.901), mixed disorders of conduct and emotions (OR, 95% CI: 0.731, 0.584-  
258 0.916), mixed disorders of conduct and emotions (OR, 95% CI: 0.664, 0.455-0.970); in addition, there  
259 were 2 endocrine system disorders and 1 immune disorder: nontoxic diffuse goitre (OR, 95% CI:  
260 1.270, 1.051-1.536), thyrotoxicosis with diffuse goitre (OR, 95% CI: 1.121, 1.017-1.235), and  
261 systemic lupus erythematosus (OR, 95% CI: 0.789, 0.643-0.968). SME, WME, and MR-Egger  
262 provided reasonably consistent associations (**Supplementary Table S5**).

263 As shown in **Figure 6**, IVW test found a statistically significant association between hospitalized  
264 COVID-19 and 24 diseases, including circulatory system (n=3): ischemic stroke (cardioembolic) [OR  
265 95%CI: 1.092 (1.004–1.186)], ischemic stroke [OR 95%CI: 1.049 (1.001–1.100)], AV-block [OR  
266 95%CI: 0.871 (0.775–0.979)]; digestive system (n=5): gastroesophageal reflux disease [OR 95%CI:  
267 1.035 (1.011–1.061)], hernia [OR 95%CI: 0.930 (0.891–0.971)], inguinal hernia [OR 95%CI: 0.915  
268 (0.869–0.965)], ulcerative colitis [OR 95%CI: 0.694 (0.588–0.819)] , Crohn's disease [OR 95%CI:  
269 0.140 (0.065–0.303)]; neoplasms (n=8): breast cancer [OR 95%CI: 1.139 (1.011–1.283)] , malignant  
270 neoplasm of breast (HER-positive, all cancers excluded) [OR 95%CI: 1.124 (1.028–1.229)],  
271 malignant neoplasm of breast (HER-positive) [OR 95%CI: 1.121 (1.027–1.224)], malignant  
272 neoplasm of breast (HER2-negative, all cancers excluded) [OR 95%CI: 1.104 (1.028–1.185)],

273 malignant neoplasm of breast (HER2-negative) [OR 95%CI: 1.100 (1.026–1.179)], malignant  
274 neoplasm of breast (all cancers excluded) [OR 95%CI: 1.097 (1.026–1.174)], pan cancer [OR 95%CI:  
275 1.003 (1.000–1.006)], oesophageal cancer [OR 95%CI: 1.001 (1.000–1.001)].

276 As shown in **Figure 7**, the IVW test found a statistically significant association between  
277 susceptibility to COVID-19 and 24 diseases, including circulatory system (n=12): deep vein  
278 thrombosis (DVT) of lower extremities [OR 95%CI: 2.392 (1.167–4.902)], DVT of lower extremities  
279 and pulmonary embolism [OR 95%CI: 2.052 (1.181–3.564)], venous thromboembolism [OR 95%CI:  
280 1.962 (1.115–3.453)], pulmonary embolism [OR 95%CI: 1.867 (1.161–3.002)], pulmonary heart  
281 disease, diseases of pulmonary circulation [OR 95%CI: 1.767 (1.142–2.733)], ischemic stroke (large  
282 artery atherosclerosis) [OR 95%CI: 1.405 (1.025–1.927)], coronary heart disease [OR 95%CI: 1.350  
283 (1.006–1.812)], myocardial infarction [OR 95%CI: 1.235 (1.012–1.509)], varicose veins [OR 95%CI:  
284 1.215 (1.001–1.474)], all-cause heart failure [OR 95%CI: 1.140 (1.009–1.287)], conduction  
285 disorders [OR 95%CI: 0.780 (0.609–0.997)], AV-block [OR 95%CI: 0.651 (0.467–0.908)]. For full  
286 results of the two-sample MR see **Supplementary Table S5**.

## 287 **4. Discussion**

288 Hypothesis-free PheWAS combined with two-sample MR analysis indicated that genetically  
289 determined COVID-19 has a detrimental effect on a broad range of diseases, in particular those  
290 impacting the circulatory system, neuropsychiatric system, neoplasms, endocrine/metabolic, and  
291 immune system. Results also highlighted the potential impact of COVID-19 on dermatologic,  
292 genitourinary, infectious, respiratory, and musculoskeletal disorders. Early detection and management  
293 of long COVID-19 could be tremendously beneficial to public health.

294 COVID-19 as an emerging infectious disease has caused a global pandemic and is spreading at  
295 a very rapid rate resulting in an extremely high proportion of people being infected. Clinical  
296 symptoms of COVID-19 range from asymptomatic infection to fatal disease [38]. The known  
297 potential mechanism is that SARS-CoV-2 enters cells via the angiotensin-converting enzyme 2 (ACE2)  
298 receptor [39]. Once inside the body, the virus undergoes replication and maturation, triggering an  
299 inflammatory response in which various cytokines are activated and infiltrate a large number of  
300 immune cells [40]. ACE2 receptors are present in multiple cell types in the body, including oral and

301 nasal mucosa, lung, heart, gastrointestinal tract, liver, kidney, spleen, brain, and arterial vein  
302 endothelial cells, partially explaining the damage caused by SARS-CoV-2 to multiple systems in the  
303 body [41]. To date, the generation of long COVID-19 has had numerous effects on the human body  
304 [3].

305 COVID-19-related cardiovascular diseases (CVDs) such as ischemic stroke, coronary heart  
306 disease, myocardial infarction, varicose veins, heart failure, and venous thromboembolism occur not  
307 only in the early stages of infection but also months later. [42] A study from the UK including 4,182  
308 community patients showed that 13.3% developed at least one persistent symptom 4 weeks after  
309 infection, half of which were thought to be of cardiac origin. [16] A recent cohort study by Yan Xie  
310 and colleagues based on the national healthcare databases from the US Department of Veterans Affairs  
311 found that COVID-19 patients had an increased risk of CVDs compared with controls even among  
312 patients who were not hospitalized during the acute phase of infection. [43] In addition, an increased  
313 risk of CVDs was observed in the presence or absence of cardiovascular risk factors or preexisting  
314 CVDs [43]. These studies suggest that COVID-19 increases the risk of CVDs after acute infection,  
315 even in individuals with a low risk of CVDs before COVID-19 infection. Our results suggest that  
316 both hospitalized and confirmed neo-coronaries may increase the risk of stroke, and therefore more  
317 attention should be paid to patients with neo-coronaries, especially hospitalized neo-coronaries, in  
318 terms of stroke prevention. For patients with COVID-19, routine screening for stroke is recommended,  
319 including clinical presentation, laboratory tests such as coagulation tests, imaging such as brain  
320 imaging, and cerebral angiography. For patients with hospitalized COVID-19, aggressive secondary  
321 prevention of stroke should be performed, and anticoagulation therapy is recommended to prevent  
322 stroke once laboratory results show a hypercoagulable state.

323 Significant associations were found between COVID-19 and the neuropsychiatric system using  
324 the PheWAS in the UK Biobank. The MR results showed that there was a strong causal effect of  
325 COVID-19 on Alzheimer's disease. Previous studies have suggested that COVID-19 may trigger  
326 clinical manifestations of neurodegenerative diseases. [44,45] Recent studies have shown that  
327 patients infected with SARS-CoV-2 exhibit reduced brain size, cognitive decline, and damage to brain  
328 regions associated with odor. [46] Previous studies have shown that SARS-CoV-2 infection activates  
329 TGF- $\beta$  signaling and oxidative overload and that neuropathological pathways normally associated

330 with AD that lead to tau hyperphosphorylation are activated in COVID-19 patients. [47] However,  
331 since most neurodegenerative diseases are late-onset and slowly progressive, current epidemiological  
332 studies may not detect such effects to an observable extent. In addition, our MR study found that  
333 COVID-19 may increase the risk of neoplasms development. However, no cancer has been reported  
334 due to the short observation period of each study. Neoplasms are a serious burden on human health  
335 care, thus, it is critical to developing active prevention and treatment solutions for at-risk populations.  
336 In addition, the impact of COVID-19 on other systems deserves further evaluation and analysis.

337 This study has several key strengths. First, a hypothesis-free PheWAS approach was used to  
338 gather evidence for a possible causal association between COVID-19 across the range of disease  
339 outcomes. Second, the availability of UK Biobank resources enabled the use of information, including  
340 hospital episode statistics and mortality data, from over 500,000 individuals. Third, the SNPs used to  
341 construct the PRS score were from the most comprehensive COVID-19 GWAS developed to this  
342 point.[34] Fourth, the comprehensive analysis strategy included several two-sample univariate MR  
343 approaches that could overcome potential pleiotropic effects and avoid overestimating results.

344 This study also had some limitations. Due to a lack of more publicly available summary-level  
345 data, the two-sample MR method could not be used to repeat all results from the PheWAS analysis.  
346 In addition, despite the large sample size, UK Biobank is not representative of the general population;  
347 [48] however, estimations of risk factor associations are generalizable.[49] Third, the analyses were  
348 restricted to participants of White-British descent, thus the generalisability of results to other ethnic  
349 groups should be made with caution.

## 350 **Conclusions**

351 This study shows a causal effect of genetically determined COVID-19 on a broad range of diseases,  
352 in particular those impacting the circulatory system, neuropsychiatric system, neoplasms, immune  
353 system, and digestive systems. Early detection and management of long COVID-19 could be a  
354 tremendous benefit to public health.

---

## 355 **References**

- 356 1. None TL. Facing up to long COVID. *Lancet* (London, England) [Internet]. *Lancet*; 2020 [cited  
357 2023 Feb 1];396. Available from: <https://pubmed.ncbi.nlm.nih.gov/33308453/>
- 358 2. Meeting the challenge of long COVID. *Nature medicine* [Internet]. *Nat Med*; 2020 [cited 2023 Feb  
359 1];26. Available from: <https://pubmed.ncbi.nlm.nih.gov/33288947/>
- 360 3. The Lancet null. Understanding long COVID: a modern medical challenge. *Lancet*. 2021;398:725.
- 361 4. NIH experts discuss post-acute COVID-19 [Internet]. National Institutes of Health (NIH). 2021  
362 [cited 2023 Feb 1]. Available from: [https://www.nih.gov/news-events/news-releases/nih-experts-](https://www.nih.gov/news-events/news-releases/nih-experts-discuss-post-acute-covid-19)  
363 [discuss-post-acute-covid-19](https://www.nih.gov/news-events/news-releases/nih-experts-discuss-post-acute-covid-19)
- 364 5. Expanding our understanding of post COVID-19 condition: report of a WHO webinar - 9 February  
365 2021 [Internet]. [cited 2023 Feb 1]. Available from:  
366 <https://www.who.int/publications/i/item/9789240025035>
- 367 6. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV, WHO Clinical Case Definition Working  
368 Group on Post-COVID-19 Condition. A clinical case definition of post-COVID-19 condition by a  
369 Delphi consensus. *Lancet Infect Dis*. 2022;22:e102–7.
- 370 7. Overview | COVID-19 rapid guideline: managing COVID-19 | Guidance | NICE [Internet]. [cited  
371 2023 Feb 1]. Available from: <https://www.nice.org.uk/guidance/NG191>
- 372 8. Datta SD, Talwar A, Lee JT. A Proposed Framework and Timeline of the Spectrum of Disease Due  
373 to SARS-CoV-2 Infection: Illness Beyond Acute Infection and Public Health Implications. *JAMA*.  
374 2020;324:2251–2.
- 375 9. Crispo A, Bimonte S, Porciello G, Forte CA, Cuomo G, Montagnese C, et al. Strategies to evaluate  
376 outcomes in long-COVID-19 and post-COVID survivors. *Infect Agent Cancer*. 2021;16:62.
- 377 10. C H, L H, Y W, X L, L R, X G, et al. 6-month consequences of COVID-19 in patients discharged  
378 from hospital: a cohort study. *Lancet* (London, England) [Internet]. *Lancet*; 2021 [cited 2023 Feb  
379 1];397. Available from: <https://pubmed.ncbi.nlm.nih.gov/33428867/>
- 380 11. Munblit D, Bobkova P, Spiridonova E, Shikhaleva A, Gamirova A, Blyuss O, et al. Incidence and  
381 risk factors for persistent symptoms in adults previously hospitalized for COVID-19. *Clin Exp*  
382 *Allergy*. 2021;51:1107–20.
- 383 12. Sigfrid L, Drake TM, Pauley E, Jesudason EC, Olliaro P, Lim WS, et al. Long Covid in adults  
384 discharged from UK hospitals after Covid-19: A prospective, multicentre cohort study using the  
385 ISARIC WHO Clinical Characterisation Protocol. *Lancet Reg Health Eur*. 2021;8:100186.
- 386 13. Huang L, Yao Q, Gu X, Wang Q, Ren L, Wang Y, et al. 1-year outcomes in hospital survivors with  
387 COVID-19: a longitudinal cohort study. *Lancet*. 2021;398:747–58.

- 
- 388 14. Méndez R, Latorre A, González-Jiménez P, Feced L, Bouzas L, Yépez K, et al. Reduced Diffusion  
389 Capacity in COVID-19 Survivors. *Ann Am Thorac Soc.* 2021;18:1253–5.
- 390 15. Townsend L, Dowds J, O’Brien K, Sheill G, Dyer AH, O’Kelly B, et al. Persistent Poor Health  
391 after COVID-19 Is Not Associated with Respiratory Complications or Initial Disease Severity. *Ann*  
392 *Am Thorac Soc.* 2021;18:997–1003.
- 393 16. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and  
394 predictors of long COVID. *Nat Med.* 2021;27:626–31.
- 395 17. Raman B, Cassar MP, Tunnicliffe EM, Filippini N, Griffanti L, Alfaro-Almagro F, et al. Medium-  
396 term effects of SARS-CoV-2 infection on multiple vital organs, exercise capacity, cognition, quality  
397 of life and mental health, post-hospital discharge. *EClinicalMedicine.* 2021;31:100683.
- 398 18. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence,  
399 and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors  
400 of COVID-19. *PLoS Med.* 2021;18:e1003773.
- 401 19. Smith GD, Ebrahim S. “Mendelian randomization”: can genetic epidemiology contribute to  
402 understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32:1–22.
- 403 20. Huang JY, Labrecque JA. From GWAS to PheWAS: the search for causality in big data. *Lancet*  
404 *Digit Health.* 2019;1:e101–3.
- 405 21. Hyppönen E, Mulugeta A, Zhou A, Santhanakrishnan VK. A data-driven approach for studying  
406 the role of body mass in multiple diseases: a phenome-wide registry-based case-control study in the  
407 UK Biobank. *Lancet Digit Health.* 2019;1:e116–26.
- 408 22. Li X, Meng X, He Y, Spiliopoulou A, Timofeeva M, Wei W-Q, et al. Genetically determined  
409 serum urate levels and cardiovascular and other diseases in UK Biobank cohort: A phenome-wide  
410 mendelian randomization study. *PLoS Med.* 2019;16:e1002937.
- 411 23. Magnus MC, Guyatt AL, Lawn RB, Wyss AB, Trajanoska K, Küpers LK, et al. Identifying  
412 potential causal effects of age at menarche: a Mendelian randomization phenome-wide association  
413 study. *BMC Med.* 2020;18:71.
- 414 24. King C, Mulugeta A, Nabi F, Walton R, Zhou A, Hyppönen E. Mendelian randomization case-  
415 control PheWAS in UK Biobank shows evidence of causality for smoking intensity in 28 distinct  
416 clinical conditions. *EClinicalMedicine.* 2020;26:100488.
- 417 25. Li S, Schooling CM. A phenome-wide association study of ABO blood groups. *BMC Med.*  
418 2020;18:334.
- 419 26. Si S, Li J, Tewara MA, Xue F. Genetically Determined Chronic Low-Grade Inflammation and  
420 Hundreds of Health Outcomes in the UK Biobank and the FinnGen Population: A Phenome-Wide  
421 Mendelian Randomization Study. *Front Immunol.* 2021;12:720876.



- 
- 422 27. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access  
423 resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS*  
424 *Med.* 2015;12:e1001779.
- 425 28. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource  
426 with deep phenotyping and genomic data. *Nature.* 2018;562:203–9.
- 427 29. COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative  
428 to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus  
429 pandemic. *Eur J Hum Genet.* 2020;28:715–8.
- 430 30. Yuan S, Mason AM, Burgess S, Larsson SC. Genetic liability to insomnia in relation to  
431 cardiovascular diseases: a Mendelian randomisation study. *Eur J Epidemiol.* 2021;36:393–400.
- 432 31. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic  
433 comparison of phenome-wide association study of electronic medical record data and genome-  
434 wide association study data. *Nat Biotechnol.* 2013;31:1102–10.
- 435 32. Li X, Meng X, Spiliopoulou A, Timofeeva M, Wei W-Q, Gifford A, et al. MR-PheWAS: exploring  
436 the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK  
437 Biobank. *Ann Rheum Dis.* 2018;77:1039–47.
- 438 33. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-  
439 wide association studies in the R environment. *Bioinformatics.* 2014;30:2375–6.
- 440 34. Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, et al. Genome-wide  
441 analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat*  
442 *Genet.* 2019;51:394–403.
- 443 35. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple  
444 genetic variants using summarized data. *Genet Epidemiol.* 2013;37:658–65.
- 445 36. J B, G DS, Pc H, S B. Consistent Estimation in Mendelian Randomization with Some Invalid  
446 Instruments Using a Weighted Median Estimator. *Genetic epidemiology [Internet]. Genet Epidemiol;*  
447 *2016 [cited 2023 Feb 15];40.* Available from: <https://pubmed.ncbi.nlm.nih.gov/27061298/>
- 448 37. J B, G DS, S B. Mendelian randomization with invalid instruments: effect estimation and bias  
449 detection through Egger regression. *International journal of epidemiology [Internet]. Int J Epidemiol;*  
450 *2015 [cited 2023 Feb 15];44.* Available from: <https://pubmed.ncbi.nlm.nih.gov/26050253/>
- 451 38. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased  
452 patients with coronavirus disease 2019: retrospective study. *BMJ.* 2020;368:m1091.
- 453 39. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-  
454 2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease  
455 Inhibitor. *Cell.* 2020;181:271-280.e8.

- 
- 456 40. Hussman JP. Cellular and Molecular Pathways of COVID-19 and Potential Points of Therapeutic  
457 Intervention. *Front Pharmacol.* 2020;11:1169.
- 458 41. Hamming I, Timens W, Bulthuis MLC, Lely AT, Navis GJ, van Goor H. Tissue distribution of  
459 ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS  
460 pathogenesis. *J Pathol.* 2004;203:631–7.
- 461 42. Raman B, Bluemke DA, Lüscher TF, Neubauer S. Long COVID: post-acute sequelae of COVID-  
462 19 with a cardiovascular focus. *Eur Heart J.* 2022;43:1157–72.
- 463 43. Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. *Nat Med.*  
464 2022;28:583–90.
- 465 44. P P, P G, Jp H, T T, Rm D, R G, et al. Neurological manifestations and COVID-19: Experiences  
466 from a tertiary care center at the Frontline. *Journal of the neurological sciences [Internet]. J Neurol*  
467 *Sci*; 2020 [cited 2023 Feb 15];415. Available from: <https://pubmed.ncbi.nlm.nih.gov/32570113/>
- 468 45. Mao L, Jin H, Wang M, Hu Y, Chen S, He Q, et al. Neurologic Manifestations of Hospitalized  
469 Patients With Coronavirus Disease 2019 in Wuhan, China. *JAMA Neurol.* 2020;77:683–90.
- 470 46. Douaud G, Lee S, Alfaro-Almagro F, Arthofer C, Wang C, McCarthy P, et al. SARS-CoV-2 is  
471 associated with changes in brain structure in UK Biobank. *Nature.* 2022;604:697–707.
- 472 47. Reiken S, Sittenfeld L, Dridi H, Liu Y, Liu X, Marks AR. Alzheimer’s-like signaling in brains of  
473 COVID-19 patients. *Alzheimers Dement.* 2022;18:955–65.
- 474 48. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of  
475 Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the  
476 General Population. *Am J Epidemiol.* 2017;186:1026–34.
- 477 49. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK  
478 Biobank against representative, general population based studies with conventional response rates:  
479 prospective cohort study and individual participant meta-analysis. *BMJ.* 2020;368:m131.

480

481

482 **Declarations**

483 **Ethics approval and consent to participate:** UK Biobank received approval from the UK Biobank  
484 Research Ethics Committee (REC; REC reference 11/NW/0382). The ethical approval of summary-  
485 level statistics was not applicable but has been approved in the original research.

486 **Consent for publication:** Not applicable.

487 **Availability of data and materials:** Details of summary-level data can be found in the supplementary  
488 material. Individual-level data that support the findings of this study are available from the UK  
489 Biobank (<http://www.ukbiobank.ac.uk>) (Application ID: 51470).

490 **Competing interests:** The authors declare that they have no competing interests.

491 **Funding:** This work was supported by the National Natural Science Foundation of China (81773547  
492 and 82173625) and National Key Research and Development Program (2020YFC2003500). The  
493 corresponding author Fuzhong Xue obtained the funding. The funders had no role in this work.

494 **Authors' contributions:** FX and KZ have the conception. FX acquire the UK Biobank dataset. KZ  
495 did the statistical analyses and drafted the initial manuscript. All authors participated in the  
496 interpretation of the results, edited and reviewed the manuscript.

497 **Acknowledgment:** Thanks to UK Biobank and other original research for providing summary-level  
498 data resources.

## 499 **Figure Titles and Legends**

### 500 **Figure 1. The results of PheWAS analysis in UK Biobank.**

501 The blue line indicates the false discovery rate (FDR) threshold. The red line indicates the self-  
502 defining notable threshold ( $P_{PheWAS} < 0.01$ ) to facilitate ideal visualization. Y-axis is minus log  
503 transformed  $P$ -value of the association between COVID-19 polygenic risk score (PRS) and disease  
504 outcomes; the X-axis provides the list of labels of 17 diseases category.

### 505 **Figure 2. The odds ratios of PheWAS analysis between severe respiratory COVID-19 and** 506 **diseases.**

507 Some diseases noted with 'other' and 'un-specified' was not shown in this Figure but displayed in the  
508 supplement material; OR: odds ratio; CI: Confidence interval.

### 509 **Figure 3. The odds ratios of PheWAS analysis between hospitalized COVID-19 and diseases.**

510 Some diseases noted with 'other' and 'un-specified' was not shown in this Figure but displayed in the  
511 supplement material; OR: odds ratio; CI: Confidence interval.

### 512 **Figure 4. The odds ratios of PheWAS analysis between susceptibility COVID-19 and diseases.**

513 Some diseases noted with 'other' and 'un-specified' was not shown in this Figure but displayed in the  
514 supplement material; OR: odds ratio; CI: Confidence interval.

### 515 **Figure 5. The odds ratios of IVW analysis between severe respiratory COVID-19 and diseases.**

516 The forest plot shows a summary of statistically significant ( $P < 0.05$ ) results under IVW MR methods;  
517 IVW, inverse-variance weighted; OR: odds ratio; CI: Confidence interval;

### 518 **Figure 6. The odds ratios of IVW analysis between hospitalized COVID-19 and diseases.**

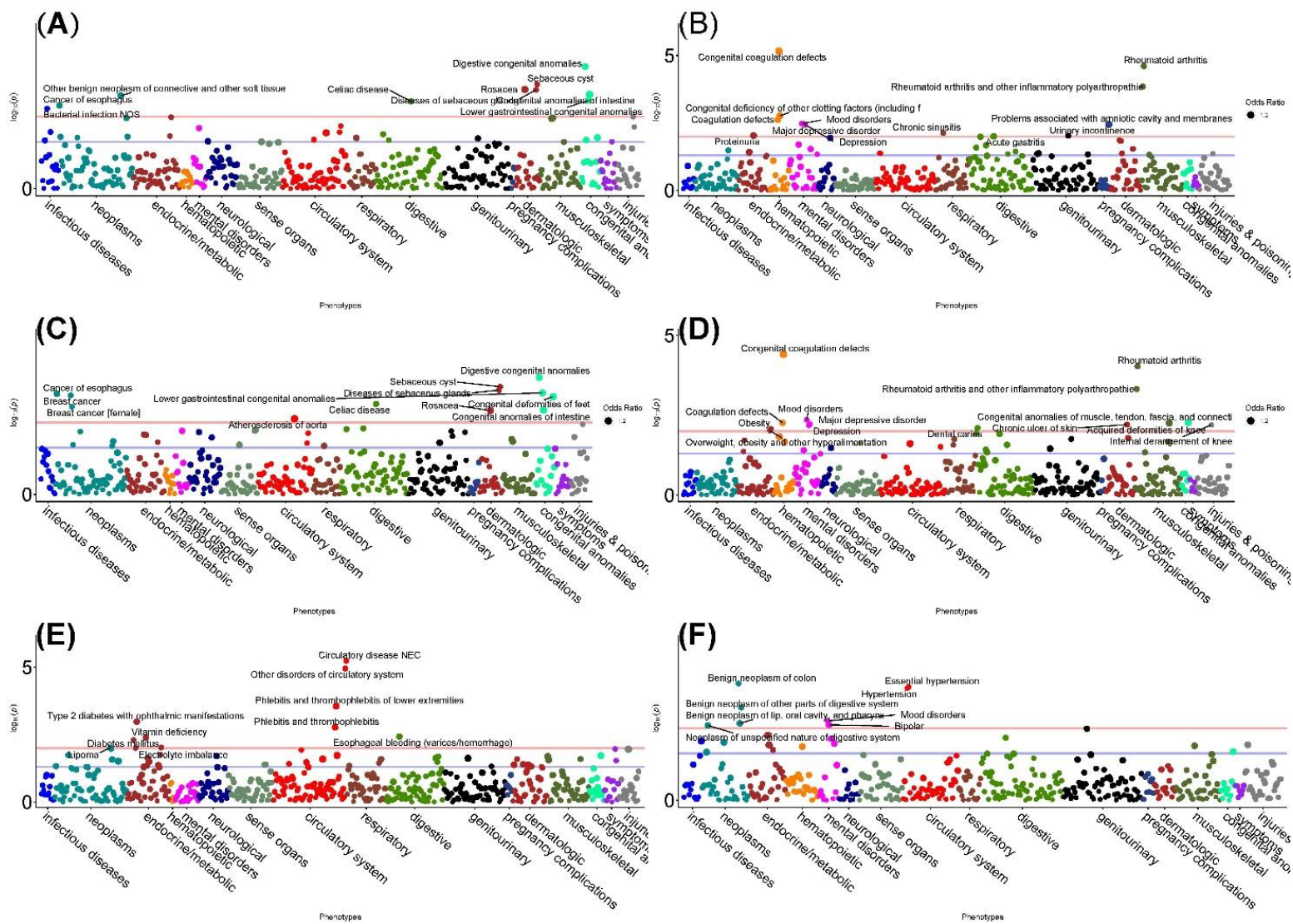
519 The forest plot shows a summary of statistically significant ( $P < 0.05$ ) results under IVW MR methods;  
520 IVW, inverse-variance weighted; OR: odds ratio; CI: Confidence interval;

### 521 **Figure 7. The odds ratios of IVW analysis between susceptibility COVID-19 and diseases**

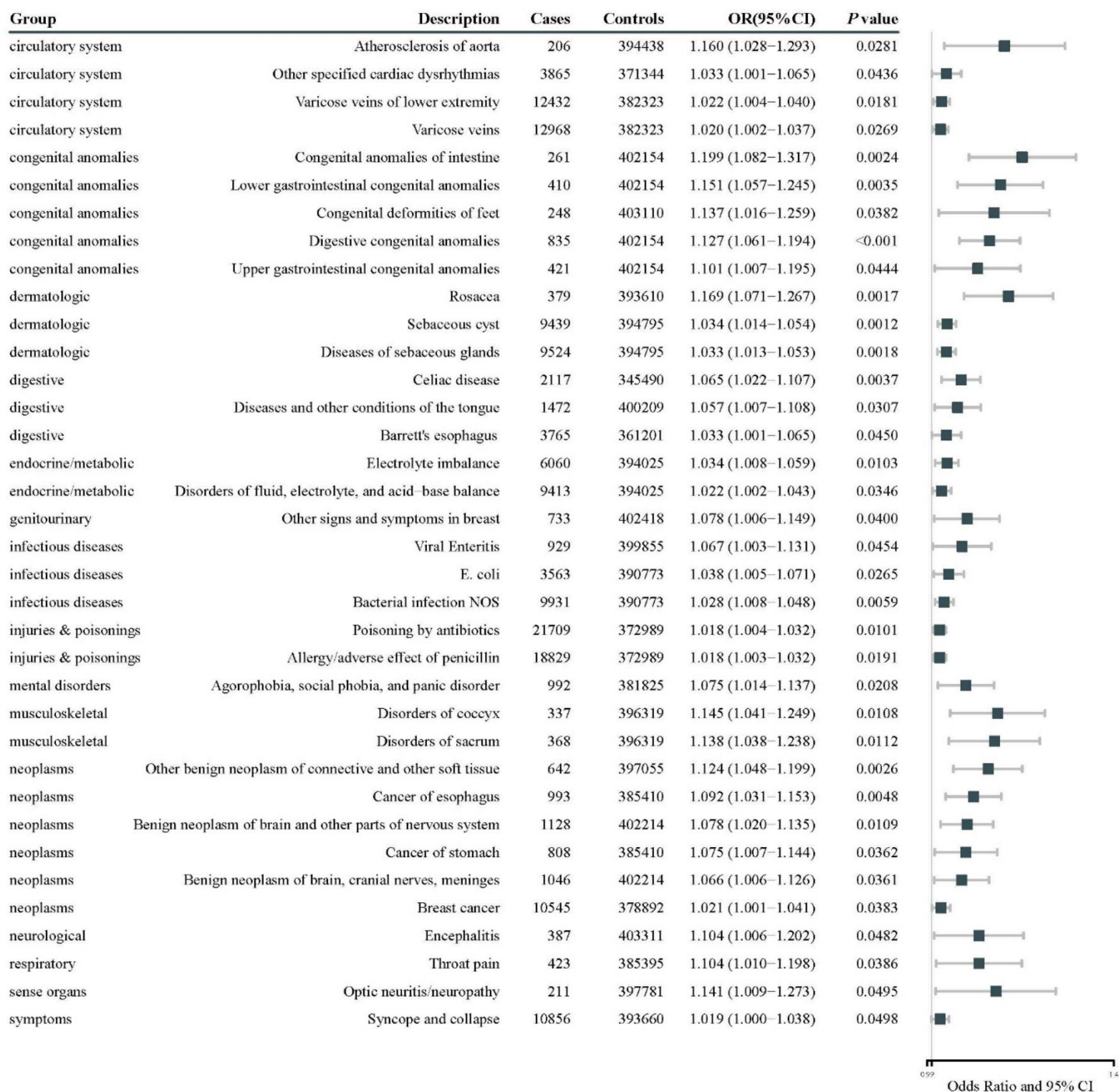
522 The forest plot shows a summary of statistically significant ( $P < 0.05$ ) results under IVW MR methods;

523 IVW, inverse-variance weighted; OR: odds ratio; CI: Confidence interval;

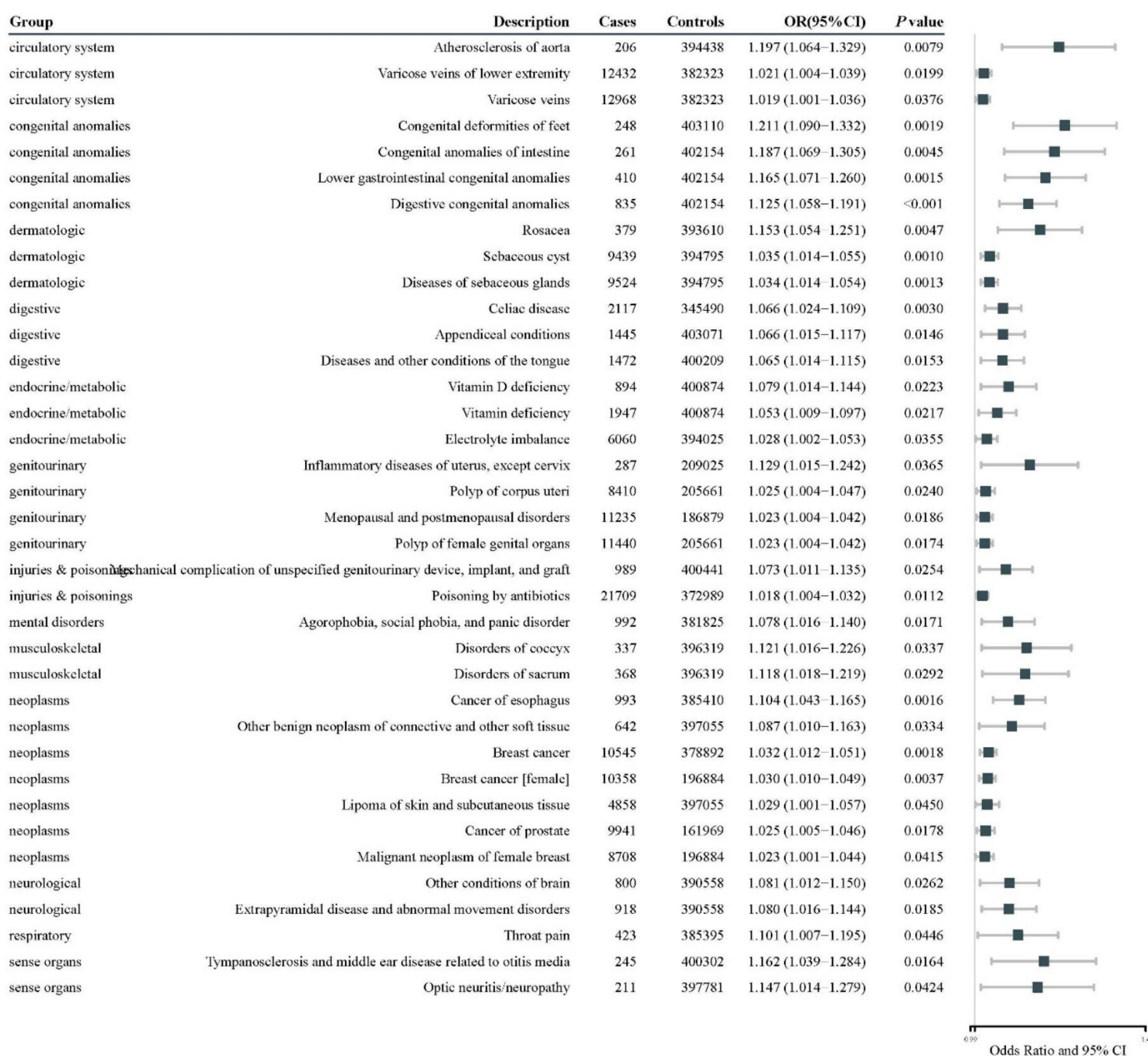
524 **Figure 1. The Manhattan plot of PheWAS analysis in UK Biobank.**



525 **Figure 2. The odds ratios of PheWAS analysis between COVID-19 and diseases.**



526 **Figure 3. The odds ratios of PheWAS analysis between COVID-19 and diseases.**

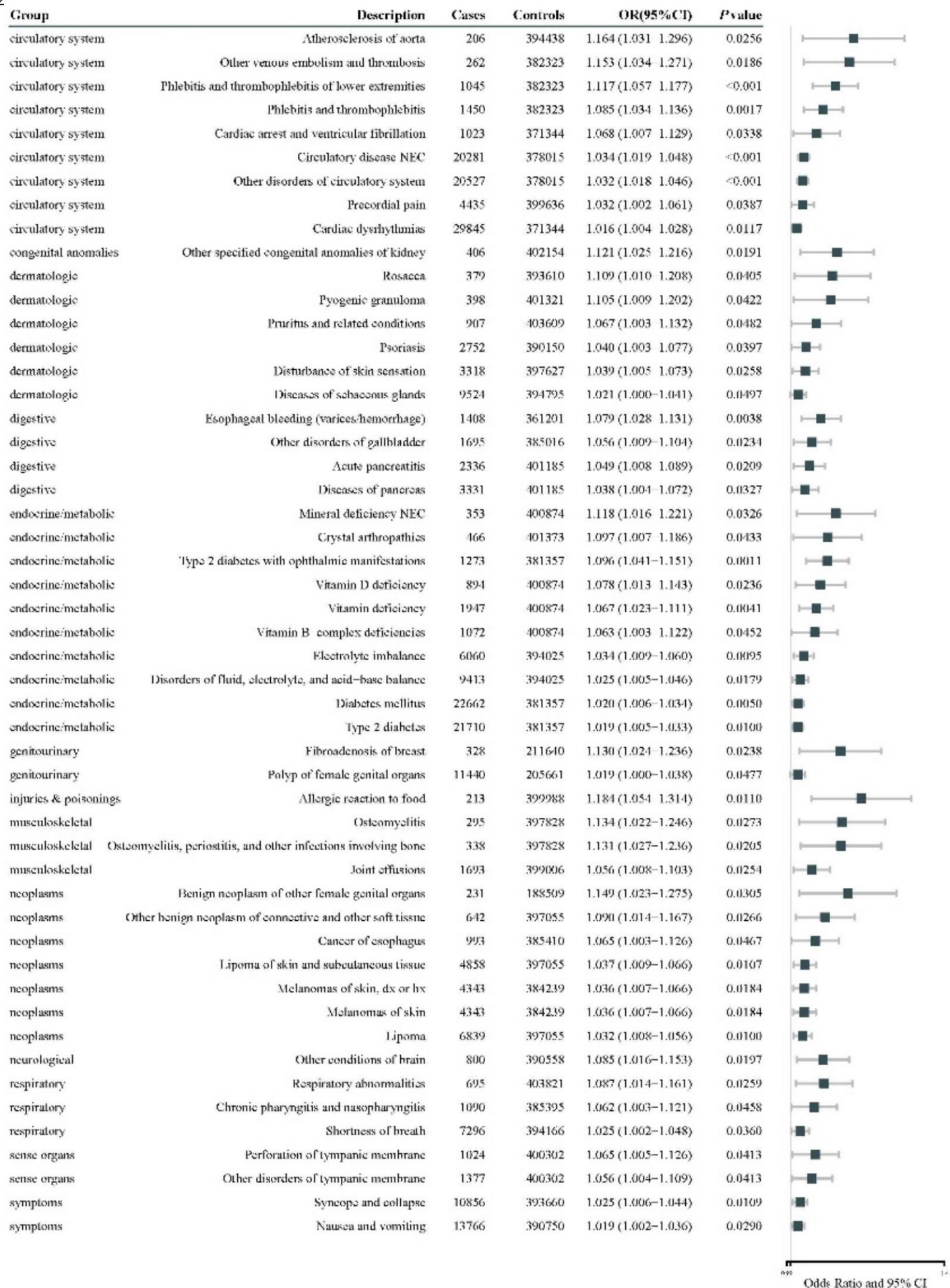


527

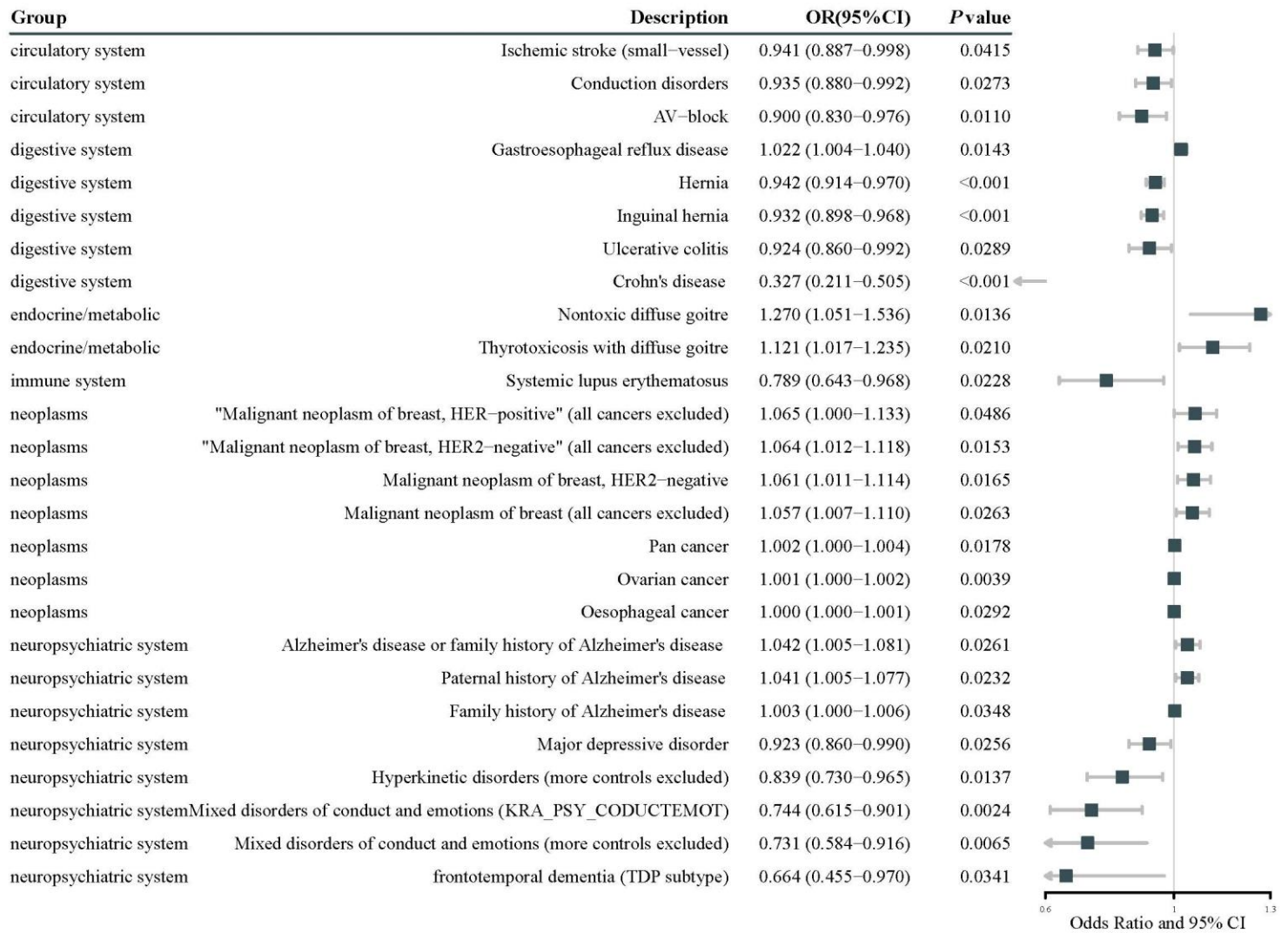


528 **Figure 4. The odds ratios of PheWAS analysis between COVID-19 and diseases.**

529

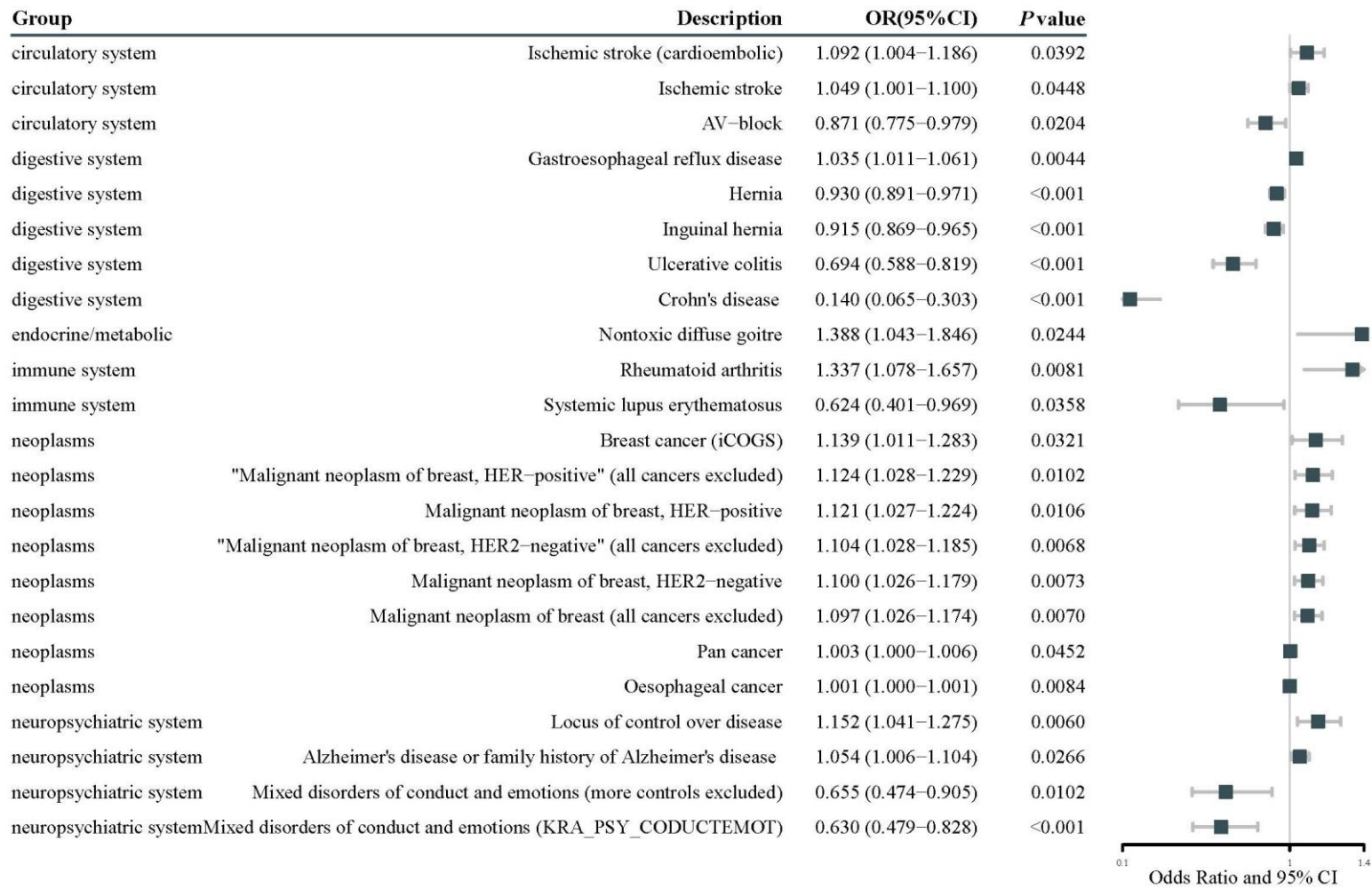


530 **Figure 5. The odds ratios of IVW analysis between COVID-19 and diseases.**

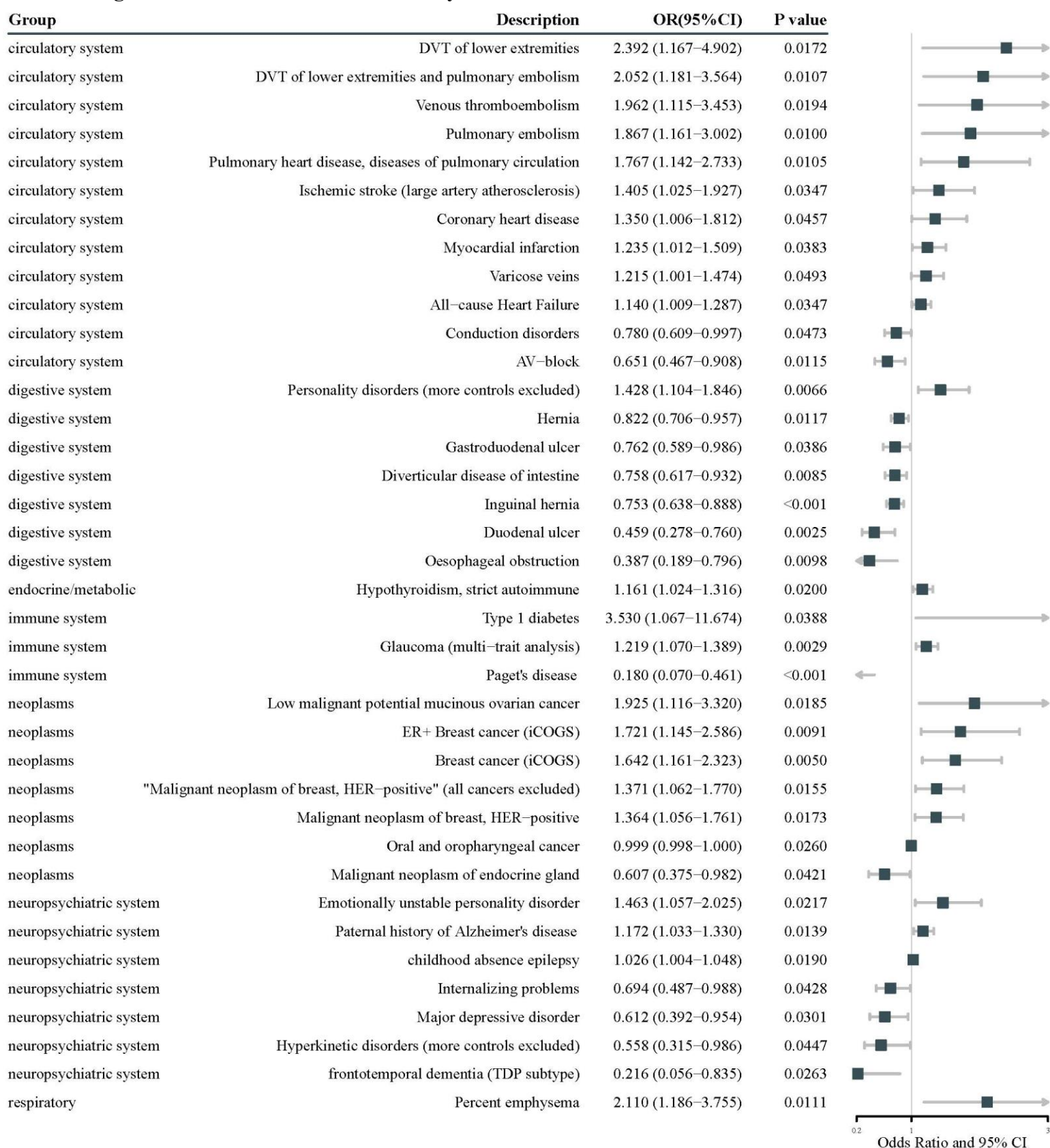


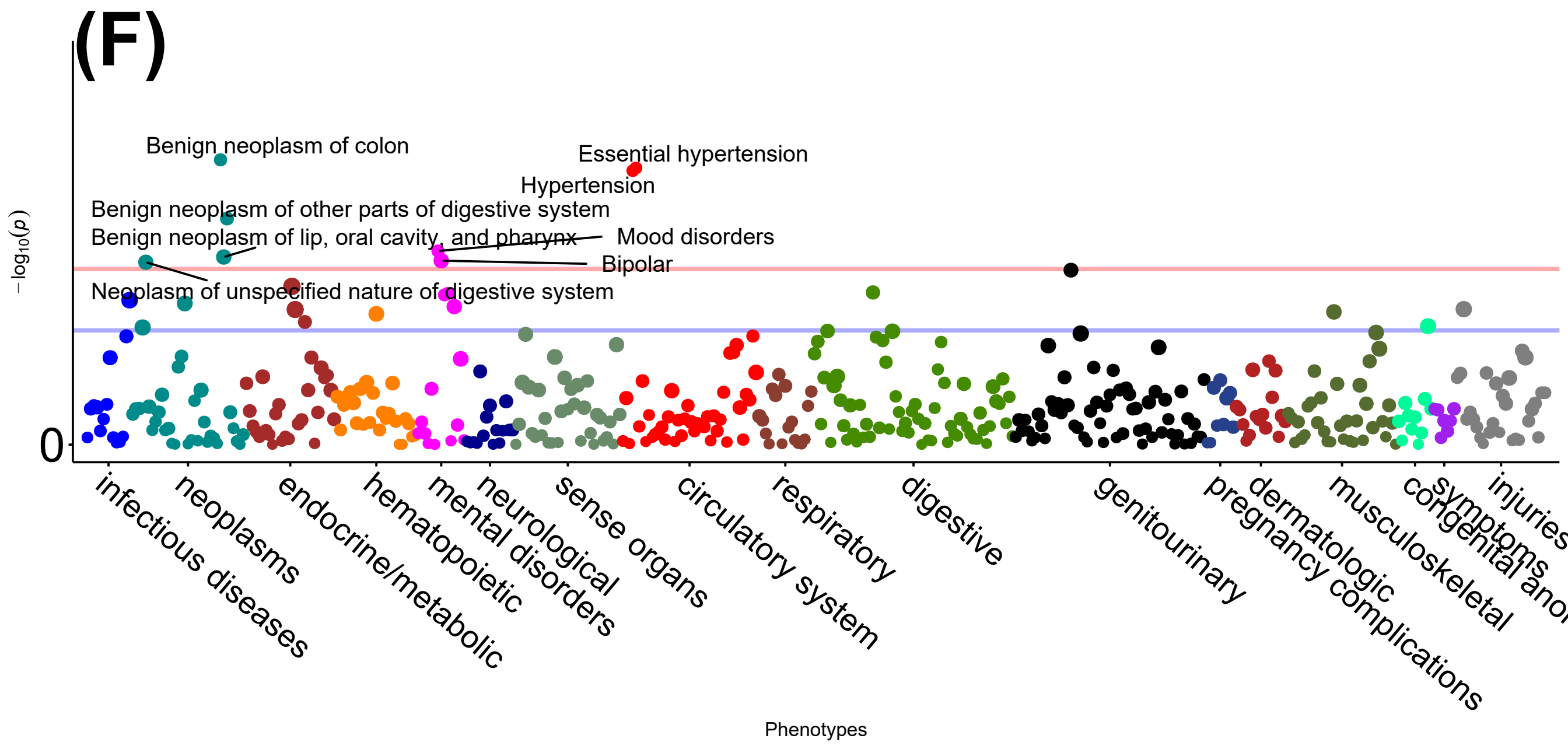
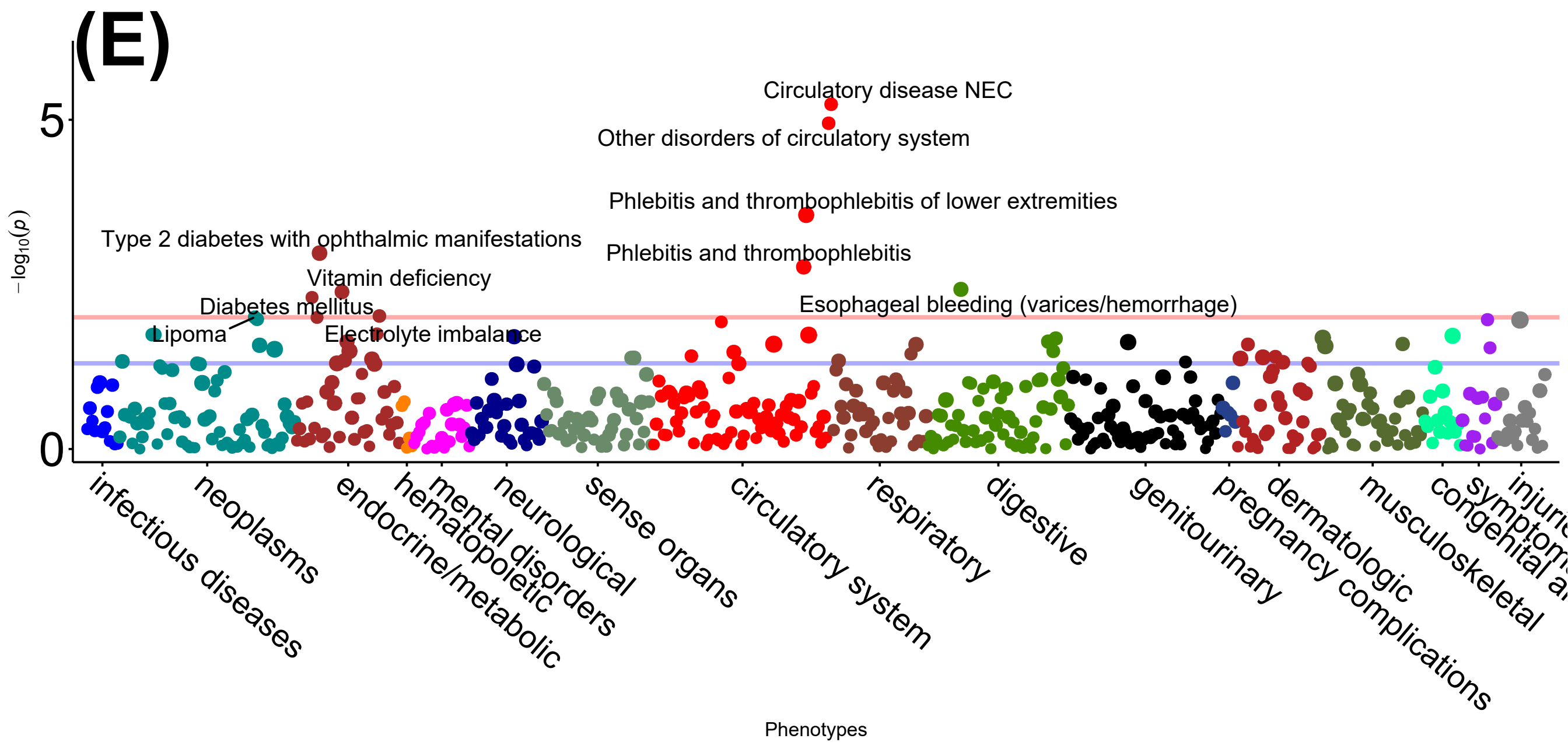
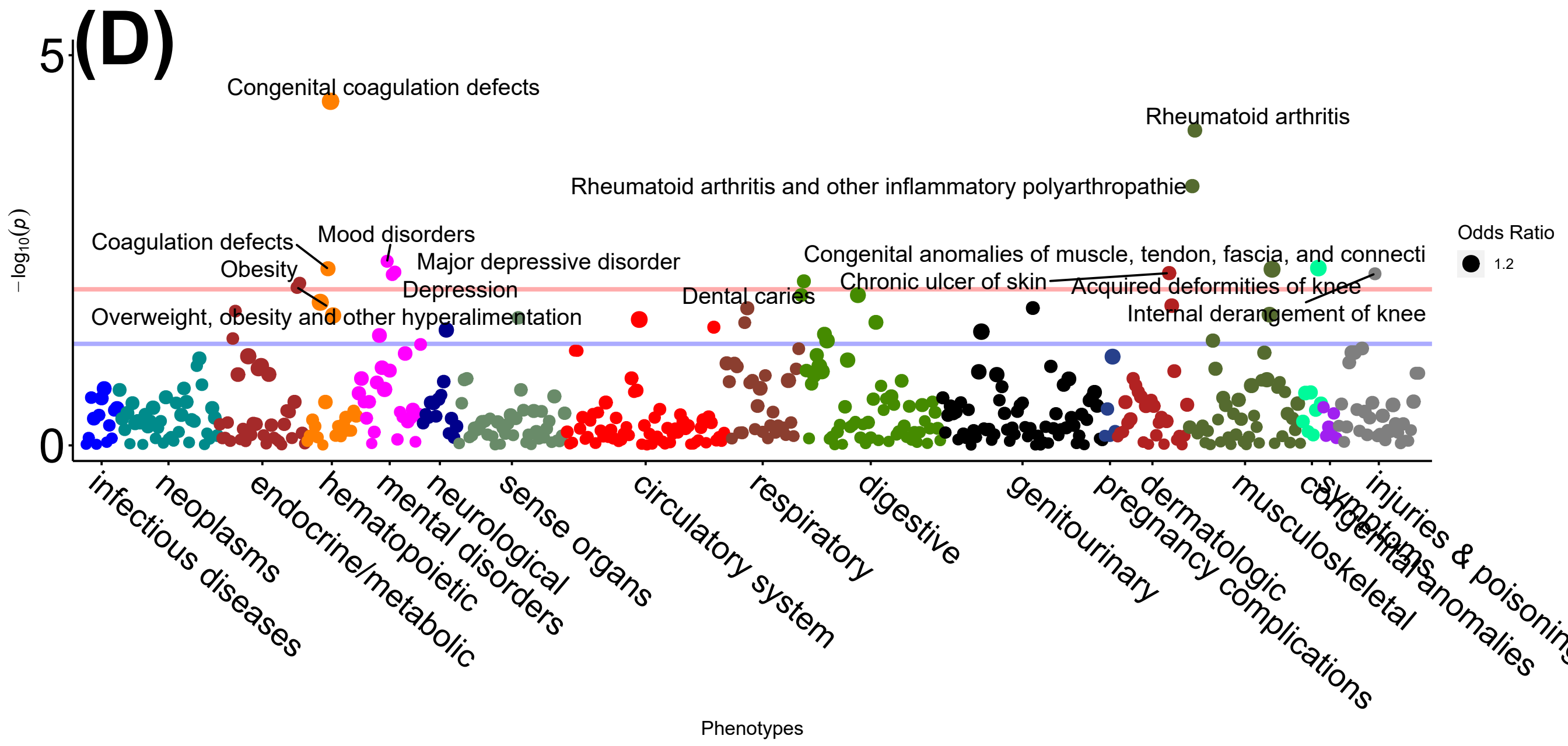
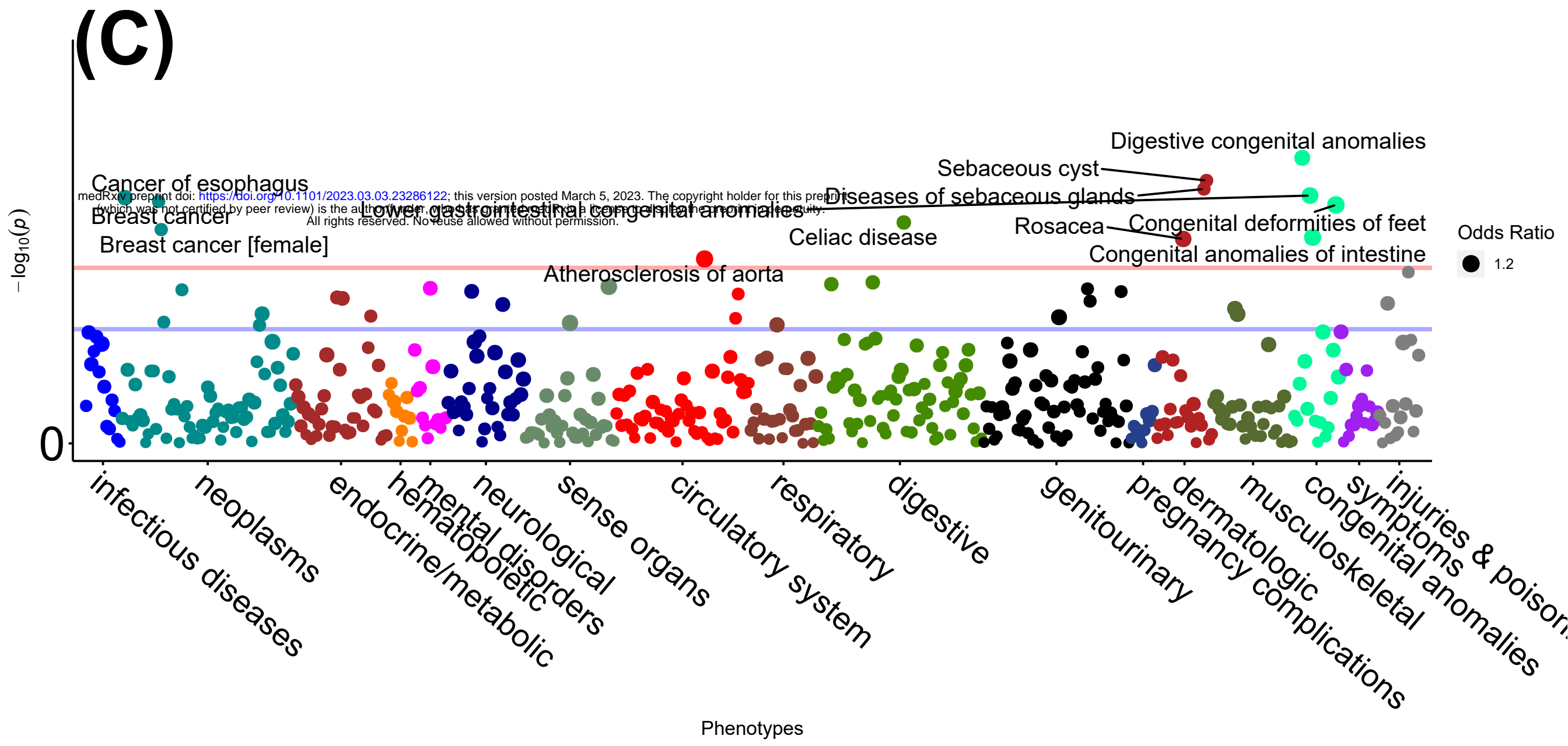
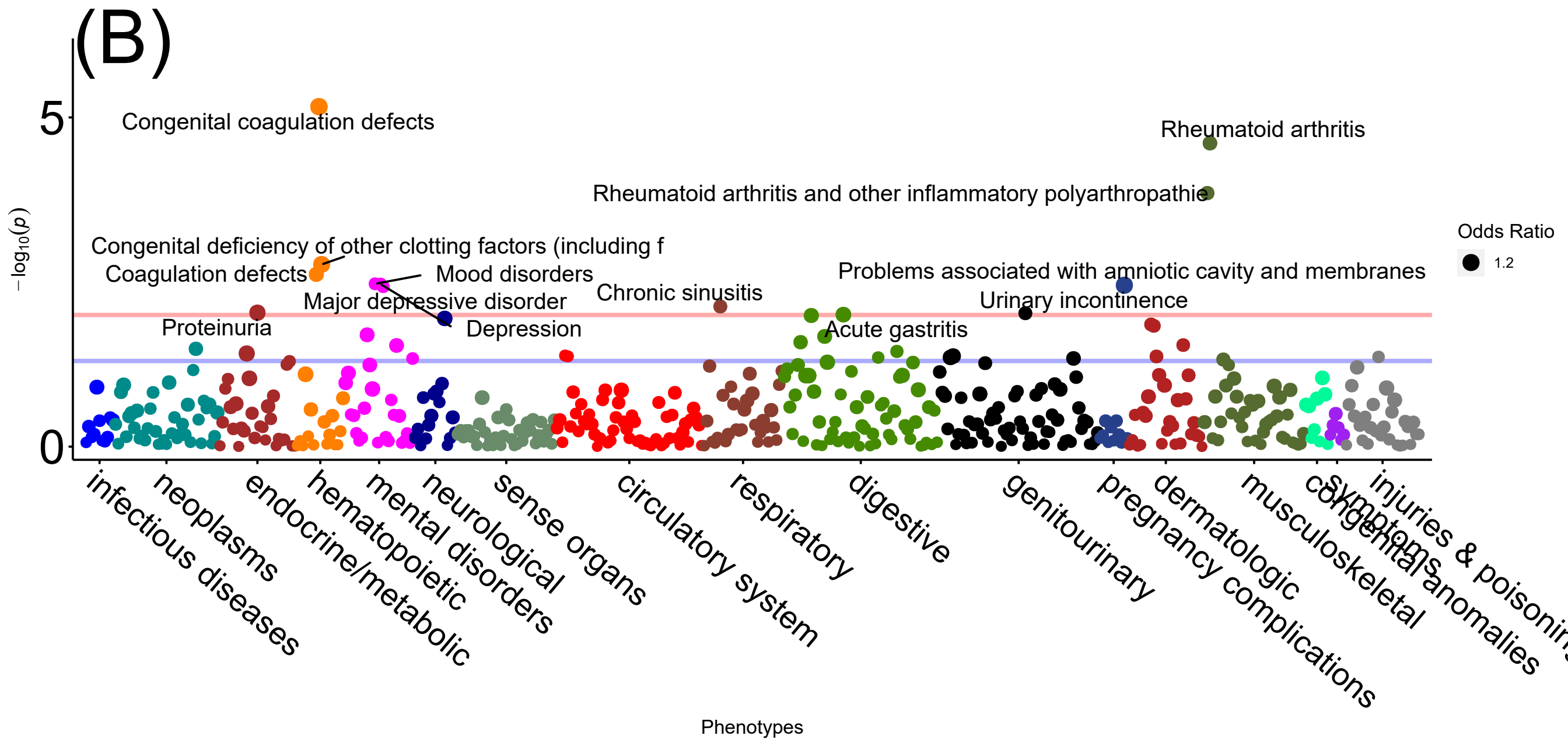
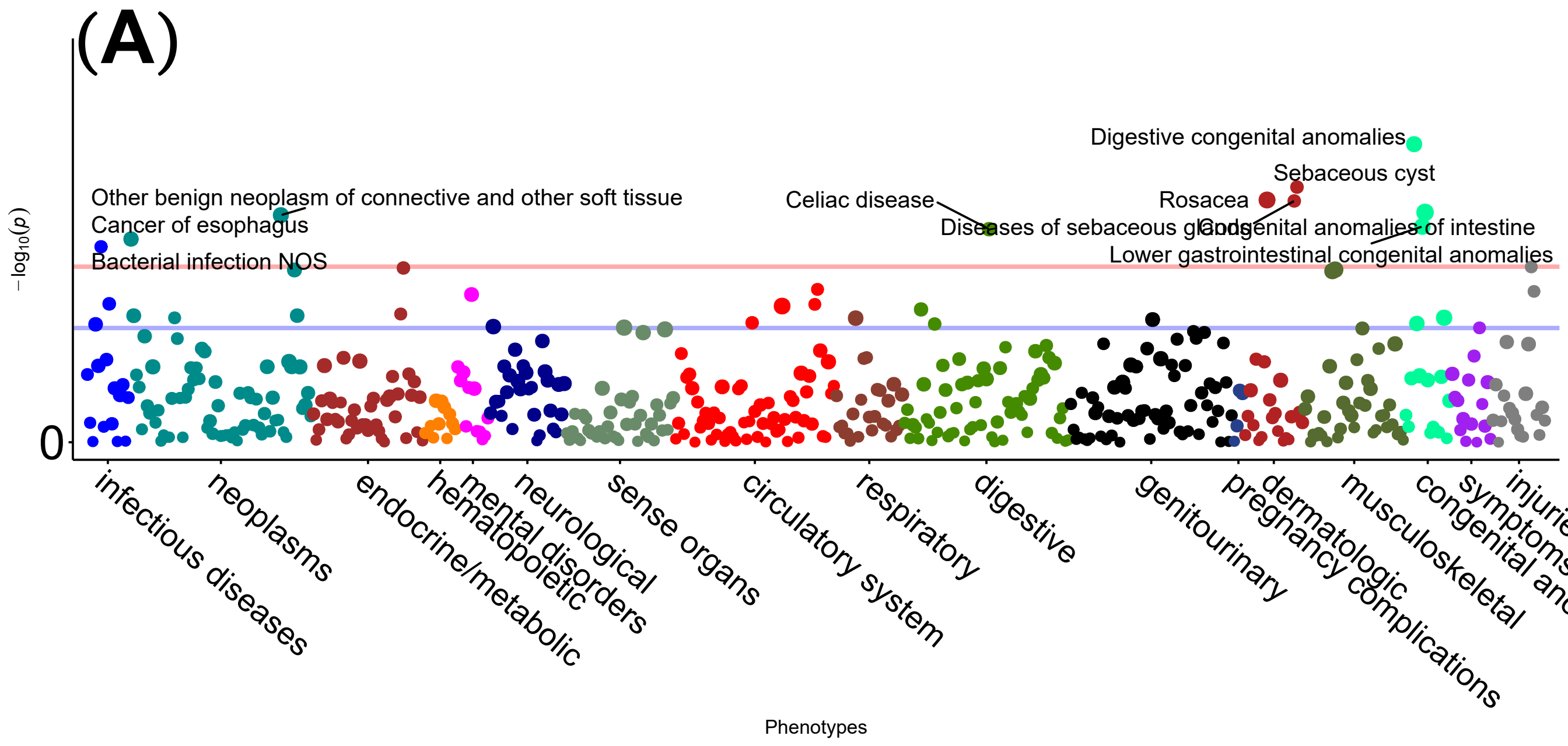
531

532 **Figure 6. The odds ratios of IVW analysis between COVID-19 and diseases.**



533 **Figure 7. The odds ratios of IVW analysis between COVID-19 and diseases.**





medRxiv preprint doi: <https://doi.org/10.1101/2023.03.03.23286122>; this version posted March 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Group	Description	Cases	Controls	OR(95%CI)	P value	
circulatory system	Atherosclerosis of aorta	906	394438	1.160 (1.028–1.293)	0.0281	
circulatory system	Other specified cardiac dysrhythmias	3865	371344	1.033 (1.001–1.065)	0.0436	
circulatory system	Varicose veins of lower extremity	12432	382323	1.022 (1.004–1.040)	0.0181	
circulatory system	Varicose veins	12968	382323	1.020 (1.002–1.037)	0.0269	
congenital anomalies	Congenital anomalies of intestine	261	402154	1.199 (1.082–1.317)	0.0024	
congenital anomalies	Lower gastrointestinal congenital anomalies	410	402154	1.151 (1.057–1.245)	0.0035	
congenital anomalies	Congenital deformities of feet	248	403110	1.137 (1.016–1.259)	0.0382	
congenital anomalies	Digestive congenital anomalies	835	402154	1.127 (1.061–1.194)	<0.001	
congenital anomalies	Upper gastrointestinal congenital anomalies	421	402154	1.101 (1.007–1.195)	0.0444	
dermatologic	Rosacea	379	393610	1.169 (1.071–1.267)	0.0017	
dermatologic	Sebaceous cyst	9439	394795	1.034 (1.014–1.054)	0.0012	
dermatologic	Diseases of sebaceous glands	9524	394795	1.033 (1.013–1.053)	0.0018	
digestive	Celiac disease	2117	345490	1.065 (1.022–1.107)	0.0037	
digestive	Diseases and other conditions of the tongue	1472	400209	1.057 (1.007–1.108)	0.0307	
digestive	Barrett's esophagus	3765	361201	1.033 (1.001–1.065)	0.0450	
endocrine/metabolic	Electrolyte imbalance	6060	394025	1.034 (1.008–1.059)	0.0103	
endocrine/metabolic	Disorders of fluid, electrolyte, and acid–base balance	9413	394025	1.022 (1.002–1.043)	0.0346	
genitourinary	Other signs and symptoms in breast	733	402418	1.078 (1.006–1.149)	0.0400	
infectious diseases	Viral Enteritis	929	399855	1.067 (1.003–1.131)	0.0454	
infectious diseases	E. coli	3563	390773	1.038 (1.005–1.071)	0.0265	
infectious diseases	Bacterial infection NOS	9931	390773	1.028 (1.008–1.048)	0.0059	
injuries & poisonings	Poisoning by antibiotics	21709	372989	1.018 (1.004–1.032)	0.0101	
injuries & poisonings	Allergy/adverse effect of penicillin	18829	372989	1.018 (1.003–1.032)	0.0191	
mental disorders	Agorophobia, social phobia, and panic disorder	992	381825	1.075 (1.014–1.137)	0.0208	
musculoskeletal	Disorders of coccyx	337	396319	1.145 (1.041–1.249)	0.0108	
musculoskeletal	Disorders of sacrum	368	396319	1.138 (1.038–1.238)	0.0112	
neoplasms	Other benign neoplasm of connective and other soft tissue	642	397055	1.124 (1.048–1.199)	0.0026	
neoplasms	Cancer of esophagus	993	385410	1.092 (1.031–1.153)	0.0048	
neoplasms	Benign neoplasm of brain and other parts of nervous system	1128	402214	1.078 (1.020–1.135)	0.0109	
neoplasms	Cancer of stomach	808	385410	1.075 (1.007–1.144)	0.0362	
neoplasms	Benign neoplasm of brain, cranial nerves, meninges	1046	402214	1.066 (1.006–1.126)	0.0361	
neoplasms	Breast cancer	10545	378892	1.021 (1.001–1.041)	0.0383	
neurological	Encephalitis	387	403311	1.104 (1.006–1.202)	0.0482	
respiratory	Throat pain	423	385395	1.104 (1.010–1.198)	0.0386	
sense organs	Optic neuritis/neuropathy	211	397781	1.141 (1.009–1.273)	0.0495	
symptoms	Syncope and collapse	10856	393660	1.019 (1.000–1.038)	0.0498	

Odds Ratio and 95% CI

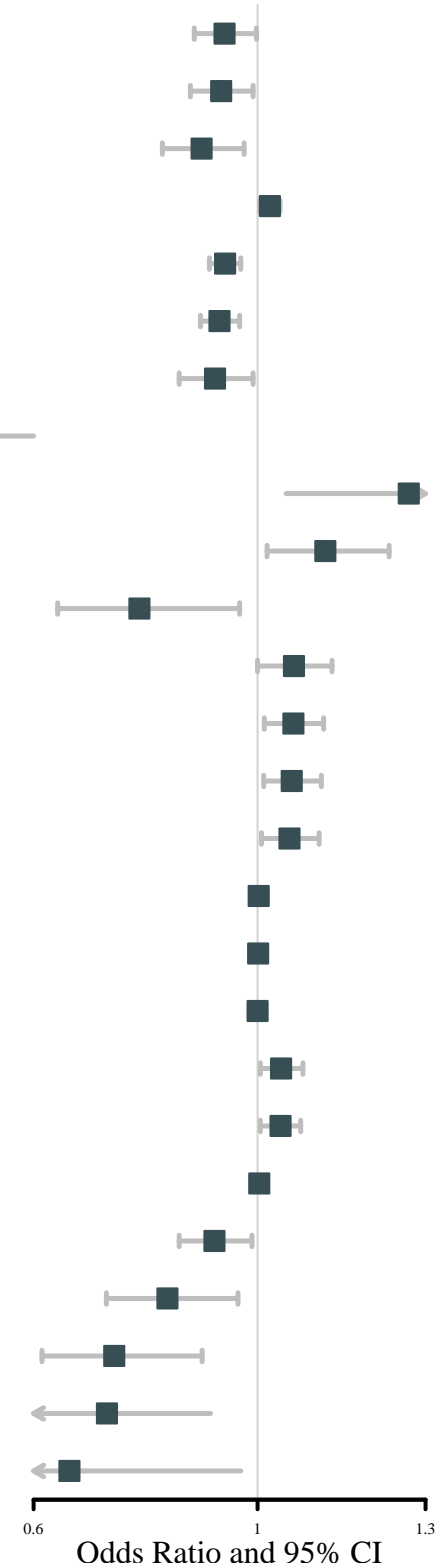
Group	Description	Cases	Controls	OR(95%CI)	P value
circulatory system	Atherosclerosis of aorta	206	394438	1.197 (1.064–1.329)	0.0079
circulatory system	Varicose veins of lower extremity	12432	382323	1.021 (1.004–1.039)	0.0199
circulatory system	Varicose veins	12968	382323	1.019 (1.001–1.036)	0.0376
congenital anomalies	Congenital deformities of feet	248	403110	1.211 (1.090–1.332)	0.0019
congenital anomalies	Congenital anomalies of intestine	261	402154	1.187 (1.069–1.305)	0.0045
congenital anomalies	Lower gastrointestinal congenital anomalies	410	402154	1.165 (1.071–1.260)	0.0015
congenital anomalies	Digestive congenital anomalies	835	402154	1.125 (1.058–1.191)	<0.001
dermatologic	Rosacea	379	393610	1.153 (1.054–1.251)	0.0047
dermatologic	Sebaceous cyst	9439	394795	1.035 (1.014–1.055)	0.0010
dermatologic	Diseases of sebaceous glands	9524	394795	1.034 (1.014–1.054)	0.0013
digestive	Celiac disease	2117	345490	1.066 (1.024–1.109)	0.0030
digestive	Appendiceal conditions	1445	403071	1.066 (1.015–1.117)	0.0146
digestive	Diseases and other conditions of the tongue	1472	400209	1.065 (1.014–1.115)	0.0153
endocrine/metabolic	Vitamin D deficiency	894	400874	1.079 (1.014–1.144)	0.0223
endocrine/metabolic	Vitamin deficiency	1947	400874	1.053 (1.009–1.097)	0.0217
endocrine/metabolic	Electrolyte imbalance	6060	394025	1.028 (1.002–1.053)	0.0355
genitourinary	Inflammatory diseases of uterus, except cervix	287	209025	1.129 (1.015–1.242)	0.0365
genitourinary	Polyp of corpus uteri	8410	205661	1.025 (1.004–1.047)	0.0240
genitourinary	Menopausal and postmenopausal disorders	11235	186879	1.023 (1.004–1.042)	0.0186
genitourinary	Polyp of female genital organs	11440	205661	1.023 (1.004–1.042)	0.0174
injuries & poisonings	Mechanical complication of unspecified genitourinary device, implant, and graft	989	400441	1.073 (1.011–1.135)	0.0254
injuries & poisonings	Poisoning by antibiotics	21709	372989	1.018 (1.004–1.032)	0.0112
mental disorders	Agoraphobia, social phobia, and panic disorder	992	381825	1.078 (1.016–1.140)	0.0171
musculoskeletal	Disorders of coccyx	337	396319	1.121 (1.016–1.226)	0.0337
musculoskeletal	Disorders of sacrum	368	396319	1.118 (1.018–1.219)	0.0292
neoplasms	Cancer of esophagus	993	385410	1.104 (1.043–1.165)	0.0016
neoplasms	Other benign neoplasm of connective and other soft tissue	642	397055	1.087 (1.010–1.163)	0.0334
neoplasms	Breast cancer	10545	378892	1.032 (1.012–1.051)	0.0018
neoplasms	Breast cancer [female]	10358	196884	1.030 (1.010–1.049)	0.0037
neoplasms	Lipoma of skin and subcutaneous tissue	4858	397055	1.029 (1.001–1.057)	0.0450
neoplasms	Cancer of prostate	9941	161969	1.025 (1.005–1.046)	0.0178
neoplasms	Malignant neoplasm of female breast	8708	196884	1.023 (1.001–1.044)	0.0415
neurological	Other conditions of brain	800	390558	1.081 (1.012–1.150)	0.0262
neurological	Extrapyramidal disease and abnormal movement disorders	918	390558	1.080 (1.016–1.144)	0.0185
respiratory	Throat pain	423	385395	1.101 (1.007–1.195)	0.0446
sense organs	Tympanosclerosis and middle ear disease related to otitis media	245	400302	1.162 (1.039–1.284)	0.0164
sense organs	Optic neuritis/neuropathy	211	397781	1.147 (1.014–1.279)	0.0424

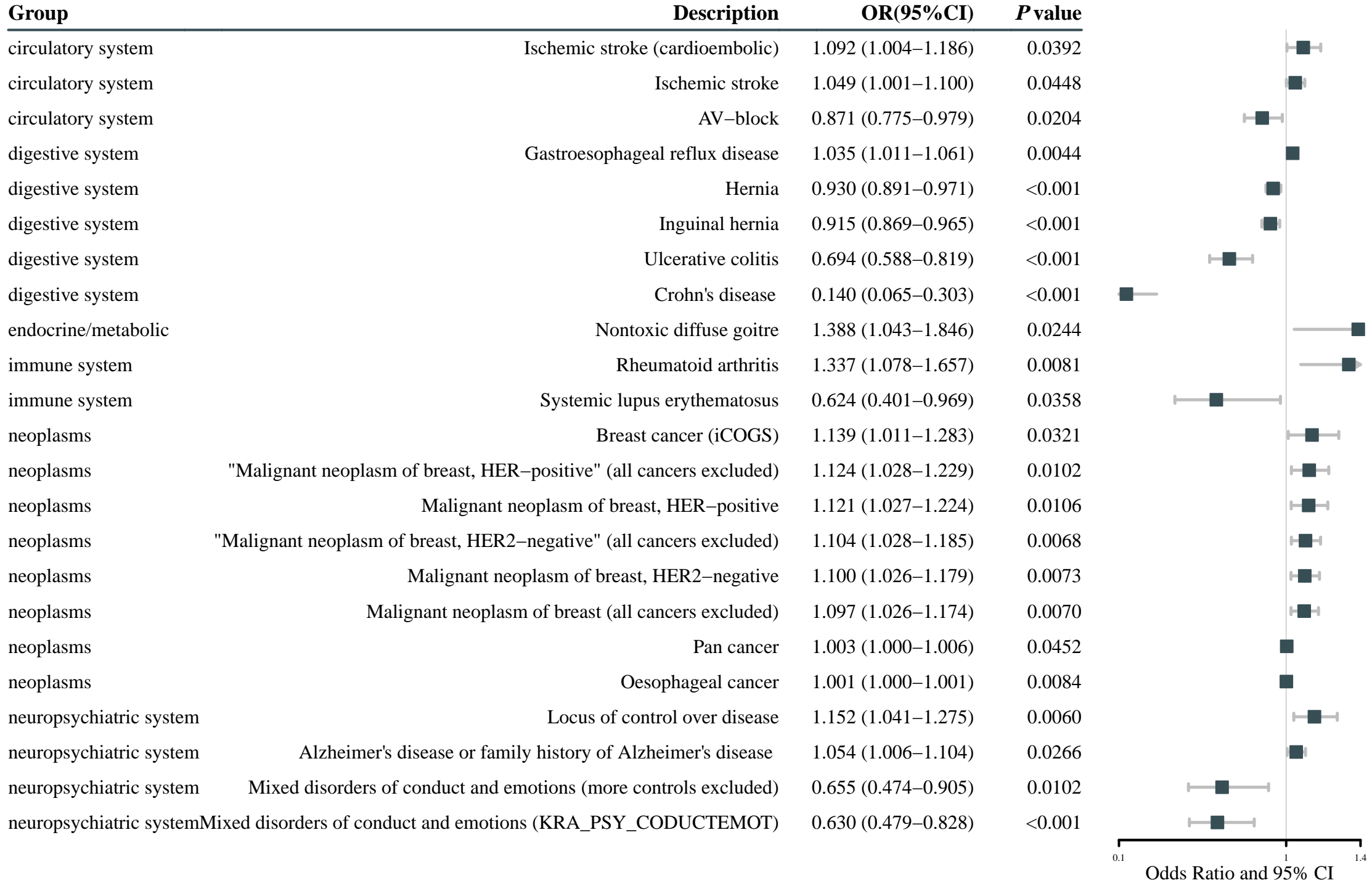
Group	Description	Cases	Controls	OR(95%CI)	P value	
circulatory system	Atherosclerosis of aorta	206	394438	1.164 (1.031–1.296)	0.0256	
circulatory system	Other venous embolism and thrombosis	262	382323	1.153 (1.034–1.271)	0.0186	
circulatory system	Phlebitis and thrombophlebitis of lower extremities	1045	382323	1.117 (1.057–1.177)	<0.001	
circulatory system	Phlebitis and thrombophlebitis	1450	382323	1.085 (1.034–1.136)	0.0017	
circulatory system	Cardiac arrest and ventricular fibrillation	1023	371344	1.068 (1.007–1.129)	0.0338	
circulatory system	Circulatory disease NEC	20281	378015	1.034 (1.019–1.048)	<0.001	
circulatory system	Other disorders of circulatory system	20527	378015	1.032 (1.018–1.046)	<0.001	
circulatory system	Precordial pain	4435	399636	1.032 (1.002–1.061)	0.0387	
circulatory system	Cardiac dysrhythmias	29845	371344	1.016 (1.004–1.028)	0.0117	
congenital anomalies	Other specified congenital anomalies of kidney	406	402154	1.121 (1.025–1.216)	0.0191	
dermatologic	Rosacea	379	393610	1.109 (1.010–1.208)	0.0405	
dermatologic	Pyogenic granuloma	398	401321	1.105 (1.009–1.202)	0.0422	
dermatologic	Pruritus and related conditions	907	403609	1.067 (1.003–1.132)	0.0482	
dermatologic	Psoriasis	2752	390150	1.040 (1.003–1.077)	0.0397	
dermatologic	Disturbance of skin sensation	3318	397627	1.039 (1.005–1.073)	0.0258	
dermatologic	Diseases of sebaceous glands	9524	394795	1.021 (1.000–1.041)	0.0497	
digestive	Esophageal bleeding (varices/hemorrhage)	1408	361201	1.079 (1.028–1.131)	0.0038	
digestive	Other disorders of gallbladder	1695	385016	1.056 (1.009–1.104)	0.0234	
digestive	Acute pancreatitis	2336	401185	1.049 (1.008–1.089)	0.0209	
digestive	Diseases of pancreas	3331	401185	1.038 (1.004–1.072)	0.0327	
endocrine/metabolic	Mineral deficiency NEC	353	400874	1.118 (1.016–1.221)	0.0326	
endocrine/metabolic	Crystal arthropathies	466	401373	1.097 (1.007–1.186)	0.0433	
endocrine/metabolic	Type 2 diabetes with ophthalmic manifestations	1273	381357	1.096 (1.041–1.151)	0.0011	
endocrine/metabolic	Vitamin D deficiency	894	400874	1.078 (1.013–1.143)	0.0236	
endocrine/metabolic	Vitamin deficiency	1947	400874	1.067 (1.023–1.111)	0.0041	
endocrine/metabolic	Vitamin B–complex deficiencies	1072	400874	1.063 (1.003–1.122)	0.0452	
endocrine/metabolic	Electrolyte imbalance	6060	394025	1.034 (1.009–1.060)	0.0095	
endocrine/metabolic	Disorders of fluid, electrolyte, and acid–base balance	9413	394025	1.025 (1.005–1.046)	0.0179	
endocrine/metabolic	Diabetes mellitus	22662	381357	1.020 (1.006–1.034)	0.0050	
endocrine/metabolic	Type 2 diabetes	21710	381357	1.019 (1.005–1.033)	0.0100	
genitourinary	Fibroadenosis of breast	328	211640	1.130 (1.024–1.236)	0.0238	
genitourinary	Polyp of female genital organs	11440	205661	1.019 (1.000–1.038)	0.0477	
injuries & poisonings	Allergic reaction to food	213	399988	1.184 (1.054–1.314)	0.0110	
musculoskeletal	Osteomyelitis	295	397828	1.134 (1.022–1.246)	0.0273	
musculoskeletal	Osteomyelitis, periostitis, and other infections involving bone	338	397828	1.131 (1.027–1.236)	0.0205	
musculoskeletal	Joint effusions	1693	399006	1.056 (1.008–1.103)	0.0254	
neoplasms	Benign neoplasm of other female genital organs	231	188509	1.149 (1.023–1.275)	0.0305	
neoplasms	Other benign neoplasm of connective and other soft tissue	642	397055	1.090 (1.014–1.167)	0.0266	
neoplasms	Cancer of esophagus	993	385410	1.065 (1.003–1.126)	0.0467	
neoplasms	Lipoma of skin and subcutaneous tissue	4858	397055	1.037 (1.009–1.066)	0.0107	
neoplasms	Melanomas of skin, dx or hx	4343	384239	1.036 (1.007–1.066)	0.0184	
neoplasms	Melanomas of skin	4343	384239	1.036 (1.007–1.066)	0.0184	
neoplasms	Lipoma	6839	397055	1.032 (1.008–1.056)	0.0100	
neurological	Other conditions of brain	800	390558	1.085 (1.016–1.153)	0.0197	
respiratory	Respiratory abnormalities	695	403821	1.087 (1.014–1.161)	0.0259	
respiratory	Chronic pharyngitis and nasopharyngitis	1090	385395	1.062 (1.003–1.121)	0.0458	
respiratory	Shortness of breath	7296	394166	1.025 (1.002–1.048)	0.0360	
sense organs	Perforation of tympanic membrane	1024	400302	1.065 (1.005–1.126)	0.0413	
sense organs	Other disorders of tympanic membrane	1377	400302	1.056 (1.004–1.109)	0.0413	
symptoms	Syncope and collapse	10856	393660	1.025 (1.006–1.044)	0.0109	
symptoms	Nausea and vomiting	13766	390750	1.019 (1.002–1.036)	0.0290	

medRxiv preprint doi: <https://doi.org/10.1101/2023.03.03.23286122>; this version posted March 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



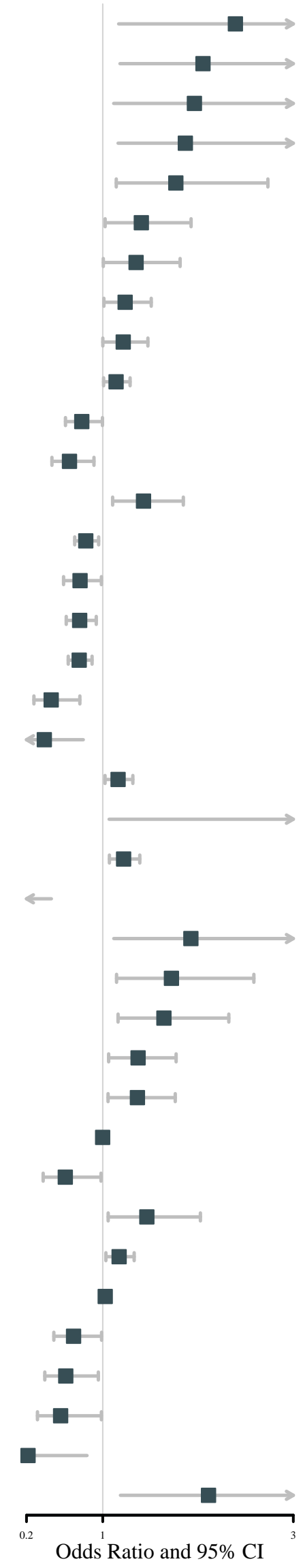
Group	Description	OR(95%CI)	P value
circulatory system	Ischemic stroke (small-vessel)	0.941 (0.887–0.998)	0.0415
circulatory system	Conduction disorders	0.935 (0.880–0.992)	0.0273
circulatory system	AV-block	0.900 (0.830–0.976)	0.0110
digestive system	Gastroesophageal reflux disease	1.022 (1.004–1.040)	0.0143
digestive system	Hernia	0.942 (0.914–0.970)	<0.001
digestive system	Inguinal hernia	0.932 (0.898–0.968)	<0.001
digestive system	Ulcerative colitis	0.924 (0.860–0.992)	0.0289
digestive system	Crohn's disease	0.327 (0.211–0.505)	<0.001 ←
endocrine/metabolic	Nontoxic diffuse goitre	1.270 (1.051–1.536)	0.0136
endocrine/metabolic	Thyrotoxicosis with diffuse goitre	1.121 (1.017–1.235)	0.0210
immune system	Systemic lupus erythematosus	0.789 (0.643–0.968)	0.0228
neoplasms	"Malignant neoplasm of breast, HER-positive" (all cancers excluded)	1.065 (1.000–1.133)	0.0486
neoplasms	"Malignant neoplasm of breast, HER2-negative" (all cancers excluded)	1.064 (1.012–1.118)	0.0153
neoplasms	Malignant neoplasm of breast, HER2-negative	1.061 (1.011–1.114)	0.0165
neoplasms	Malignant neoplasm of breast (all cancers excluded)	1.057 (1.007–1.110)	0.0263
neoplasms	Pan cancer	1.002 (1.000–1.004)	0.0178
neoplasms	Ovarian cancer	1.001 (1.000–1.002)	0.0039
neoplasms	Oesophageal cancer	1.000 (1.000–1.001)	0.0292
neuropsychiatric system	Alzheimer's disease or family history of Alzheimer's disease	1.042 (1.005–1.081)	0.0261
neuropsychiatric system	Paternal history of Alzheimer's disease	1.041 (1.005–1.077)	0.0232
neuropsychiatric system	Family history of Alzheimer's disease	1.003 (1.000–1.006)	0.0348
neuropsychiatric system	Major depressive disorder	0.923 (0.860–0.990)	0.0256
neuropsychiatric system	Hyperkinetic disorders (more controls excluded)	0.839 (0.730–0.965)	0.0137
neuropsychiatric system	Mixed disorders of conduct and emotions (KRA_PSY_CODUCTEMOT)	0.744 (0.615–0.901)	0.0024
neuropsychiatric system	Mixed disorders of conduct and emotions (more controls excluded)	0.731 (0.584–0.916)	0.0065
neuropsychiatric system	frontotemporal dementia (TDP subtype)	0.664 (0.455–0.970)	0.0341





Group	Description	OR(95%CI)	P value
circulatory system	DVT of lower extremities	2.392 (1.167–4.902)	0.0172
circulatory system	DVT of lower extremities and pulmonary embolism	2.052 (1.181–3.564)	0.0107
circulatory system	Venous thromboembolism	1.962 (1.115–3.453)	0.0194
circulatory system	Pulmonary embolism	1.867 (1.161–3.002)	0.0100
circulatory system	Pulmonary embolism, deep vein thrombosis, diseases of pulmonary circulation	1.767 (1.142–2.733)	0.0105
circulatory system	Ischemic stroke (large artery atherosclerosis)	1.405 (1.025–1.927)	0.0347
circulatory system	Coronary heart disease	1.350 (1.006–1.812)	0.0457
circulatory system	Myocardial infarction	1.235 (1.012–1.509)	0.0383
circulatory system	Varicose veins	1.215 (1.001–1.474)	0.0493
circulatory system	All-cause Heart Failure	1.140 (1.009–1.287)	0.0347
circulatory system	Conduction disorders	0.780 (0.609–0.997)	0.0473
circulatory system	AV-block	0.651 (0.467–0.908)	0.0115
digestive system	Personality disorders (more controls excluded)	1.428 (1.104–1.846)	0.0066
digestive system	Hernia	0.822 (0.706–0.957)	0.0117
digestive system	Gastroduodenal ulcer	0.762 (0.589–0.986)	0.0386
digestive system	Diverticular disease of intestine	0.758 (0.617–0.932)	0.0085
digestive system	Inguinal hernia	0.753 (0.638–0.888)	<0.001
digestive system	Duodenal ulcer	0.459 (0.278–0.760)	0.0025
digestive system	Oesophageal obstruction	0.387 (0.189–0.796)	0.0098
endocrine/metabolic	Hypothyroidism, strict autoimmune	1.161 (1.024–1.316)	0.0200
immune system	Type 1 diabetes	3.530 (1.067–11.674)	0.0388
immune system	Glaucoma (multi-trait analysis)	1.219 (1.070–1.389)	0.0029
immune system	Paget's disease	0.180 (0.070–0.461)	<0.001
neoplasms	Low malignant potential mucinous ovarian cancer	1.925 (1.116–3.320)	0.0185
neoplasms	ER+ Breast cancer (iCOGS)	1.721 (1.145–2.586)	0.0091
neoplasms	Breast cancer (iCOGS)	1.642 (1.161–2.323)	0.0050
neoplasms	"Malignant neoplasm of breast, HER-positive" (all cancers excluded)	1.371 (1.062–1.770)	0.0155
neoplasms	Malignant neoplasm of breast, HER-positive	1.364 (1.056–1.761)	0.0173
neoplasms	Oral and oropharyngeal cancer	0.999 (0.998–1.000)	0.0260
neoplasms	Malignant neoplasm of endocrine gland	0.607 (0.375–0.982)	0.0421
neuropsychiatric system	Emotionally unstable personality disorder	1.463 (1.057–2.025)	0.0217
neuropsychiatric system	Paternal history of Alzheimer's disease	1.172 (1.033–1.330)	0.0139
neuropsychiatric system	childhood absence epilepsy	1.026 (1.004–1.048)	0.0190
neuropsychiatric system	Internalizing problems	0.694 (0.487–0.988)	0.0428
neuropsychiatric system	Major depressive disorder	0.612 (0.392–0.954)	0.0301
neuropsychiatric system	Hyperkinetic disorders (more controls excluded)	0.558 (0.315–0.986)	0.0447
neuropsychiatric system	frontotemporal dementia (TDP subtype)	0.216 (0.056–0.835)	0.0263
respiratory	Percent emphysema	2.110 (1.186–3.755)	0.0111

print doi: <https://doi.org/10.1101/2023.03.03.23286122>; this version posted March 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



Odds Ratio and 95% CI