

1 **Hot Spring Residency and Disease Association:**
2 **a Crossover Gene-Environment Interaction (GxE) Study in Taiwan**

3
4 Hsin-Yu Wu^{a,b#}, Kao-Jung Chang^{a,c,d,e#}, Wei Chiu^{a,b*}, Ching-Yun Wang^{f*},
5 Yu-Tien Hsu^g, Yuan-Chih Wen^{b,h,i}, Pin-Hsuan Chiang^{a,j,k}, Yu-Hsiang Chen^{a,b},
6 He-Jhen Dai^{a,b}, Chia-Hsin Lu^{a,b}, Yi-Cheng Chen^{a,b}, Han-Ying Tsai^{a,j,k},
7 Yu-Chun Chen^{b,j,k,l,m}, Chih-Hung Hsu^{n,o}, Ai-Ru Hsieh^k, Shih-Hwa Chiou^{a,d,p},
8 Yi-Ping Yang^{a,q#}, Chih-Chien Hsu^{a,b,d+}
9 *# contributed equally, * contributed equally, + denotes the corresponding author*

10

11 ^a Department of Medical Research, Taipei Veterans General Hospital, No.201, Sec. 2,
12 Shipai Rd., Beitou District, Taipei 11217, Taiwan

13 ^b School of Medicine, National Yang Ming Chiao Tung University, No. 155, Sec. 2,
14 Linong St. Beitou Dist., Taipei 112304, Taiwan

15 ^c Institute of Clinical Medicine, National Yang Ming Chiao Tung University, No. 155,
16 Sec. 2, Linong St. Beitou Dist., Taipei 112304, Taiwan

17 ^d Department of Ophthalmology, Taipei Veterans General Hospital, No.201, Sec. 2,
18 Shipai Rd., Beitou District, Taipei 11217, Taiwan

19 ^e Department of Medical Education, Taipei Veterans General Hospital, No.201, Sec. 2,
20 Shipai Rd., Beitou District, Taipei 11217, Taiwan

21 ^f Department of Medical Education, Taichung Veterans General Hospital, No. 1650,
22 Taiwan Boulevard Sect. 4, Taichung 407219, Taiwan

23 ^g Department of Social & Behavioral Sciences, Harvard T.H. Chan School of Public
24 Health, 677 Huntington Ave, Boston, MA 02115, USA

25 ^h Department of Occupational Medicine and Clinical Toxicology, Taipei Veterans
26 General Hospital, No.201, Sec. 2, Shipai Rd., Beitou District, Taipei 11217, Taiwan

27 ⁱ Institute of Environmental and Occupational Health Sciences, National Yang Ming
28 Chiao Tung University, No. 155, Sec. 2, Linong St. Beitou Dist., Taipei 112304,
29 Taiwan.

30 ^j Big Data Center, Taipei Veterans General Hospital, No.201, Sec. 2, Shipai Rd.,
31 Beitou District, Taipei 11217, Taiwan

32 ^k Department of Statistics, Tamkang University, No.151, Yingzhuan Rd., Tamsui Dist.,
33 New Taipei 251301, Taiwan

34 ^l Institute of Hospital and Health Care Administration, National Yang Ming Chiao
35 Tung University, No. 155, Sec. 2, Linong St. Beitou Dist., Taipei 112304, Taiwan

36 ^m Department of Family Medicine, Taipei Veterans General Hospital, No.201, Sec. 2,
37 Shipai Rd., Beitou District, Taipei 11217, Taiwan

38 ⁿ Women's Hospital, The Fourth Affiliated Hospital, and Department of

39 Environmental Medicine, Zhejiang University School of Medicine, No. 866,
40 Yuhangtang Rd, Hangzhou 310058, China

41 ^o Zhejiang University School of Medicine, Hangzhou 310006 Institute of Genetics,
42 International School of Medicine, Zhejiang University, No. 866, Yuhangtang Rd,
43 Hangzhou 310058, China

44 ^p Institute of Pharmacology, National Yang Ming Chiao Tung University, No. 155, Sec.
45 2, Linong St. Beitou Dist., Taipei 112304, Taiwan

46 ^q Institute of Food Safety and Health Risk Assessment, School of Pharmaceutical
47 Sciences, National Yang Ming Chiao Tung University, No. 155, Sec. 2, Linong St.
48 Beitou Dist., Taipei 112304, Taiwan

49

50 **E-mail Address of Each Author**

51 Hsin-Yu Wu: hsinyucindy.md10@nycu.edu.tw

52 Kao-Jung Chang: michaelchang1109@gmail.com

53 Wei Chiu: alv320429@gmail.com

54 Ching-Yun Wang: wertherwang@gmail.com

55 Yu-Tien Hsu: yuh599@mail.harvard.edu

56 Yuan-Chih Wen: ck990595@gmail.com

57 Pin-Hsuan Chiang: s94120506@gmail.com

58 Yu-Hsiang Chen: ha2martinchen6050@gmail.com

59 He-Jhen Dai: janet527126@gmail.com

60 Chia-Hsin Lu: annabeth.lu.md12@nycu.edu.tw

61 Yi-Cheng Chen: cyc1214.md12@nycu.edu.tw

62 Han-Ying Tsai: hanvwww@gmail.com

63 Yu-Chun Chen: ycchen22@vghtpe.gov.tw

64 Chih-Hung Hsu: ch_hsu@zju.edu.cn

65 Ai-Ru Hsieh: airudropbox@gmail.com

66 Shih-Hwa Chiou: shchiou@vghtpe.gov.tw

67 Yi-Ping Yang: molly0103@gmail.com

68 Chih-Chien Hsu: chihchienym@gmail.com

69

70 Correspondence:

71 **Chih-Chien Hsu**, No. 201, Sec. 2, Shipai Rd., Beitou District, 11217 Taipei City,

72 Taiwan; chihchienym@gmail.com, +886-02-28757325

73

Abstract

74 **Background**

75 The advent of genetic biobanking has powered gene-environment interaction (GxE)
76 studies in various disease contexts. Therefore, we aimed to discover novel GxE
77 effects that address hot spring residency as a risk to inconspicuous disease
78 association.

79 **Methods**

80 A complete genetic and demographic registry comprising 129,451 individuals was
81 obtained from Taiwan Biobank (TWB). Geographical disease prevalence was
82 analyzed to identify putative disease association with hot-spring residency,
83 multivariable regression and logistic regression were rechecked to exclude
84 socioeconomic confounders in geographical-disease association. Genome-wide
85 association study (GWAS), gene ontology (GO), and protein-protein interaction (PPI)
86 analysis identified predisposing genetic factors among hot-spring-associated diseases.
87 Lastly, a polygenic risk score (PRS) model was formulated to stratify environmental
88 susceptibility in accord to their genetic predisposition.

89 **Results**

90 After socioeconomic covariate adjustment, prevalence of dry eye disease (DED) and
91 valvular heart disease (VHD) was significantly associated with hot spring distribution.
92 Through single nucleotide polymorphisms (SNPs) discovery and subsequent PPI
93 pathway aggregation, CDKL2 and BMP2 kinase pathways were significantly
94 enriched in hot-spring specific DED and VHD functional SNPs. Notably, PRS
95 predicted disease well in hot spring regions (PRS_{DED} : AUC=0.9168; PRS_{VHD}
96 AUC=0.8163). Hot spring and discovered SNPs contributed to crossover GxE effect
97 on both DED (relative risk $(RR)_{G+E} = 0.99$; $RR_{G+E+} = 0.35$; $RR_{G+E-} = 2.04$) and VHD
98 ($RR_{G+E} = 0.99$; $RR_{G+E+} = 0.49$; $RR_{G+E-} = 2.01$).

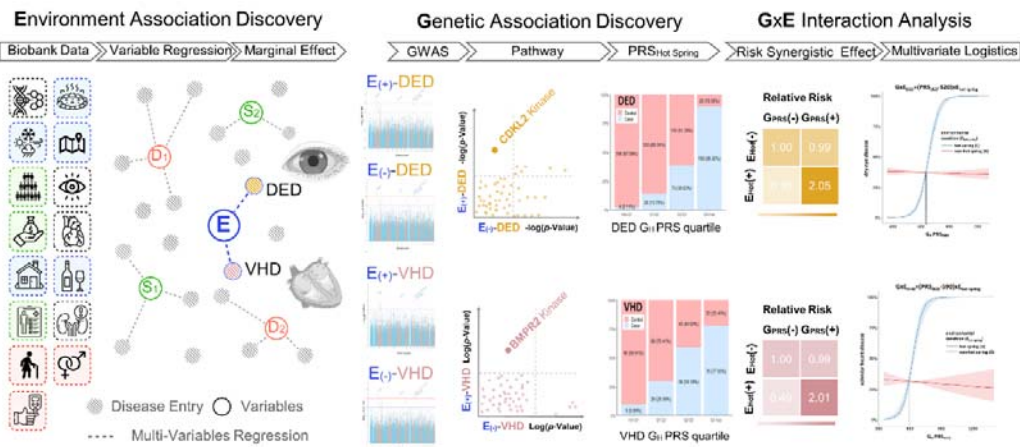
99 **Conclusion**

100 We identified hot-spring exposure as a modifiable risk in the PRS predicted GxE
101 context of DED and VHD.

102

103 **Keywords**

104 Gene-environment association, Hot spring, Dry eye disease, Valvular heart disease,
105 Genome-wide association study, Taiwan biobank



106

107 **1. Introduction**

108 Gene-environment interaction (GxE) embodies a pathogenic context in which the
109 impact of environmental exposures modifies our susceptibility to genetic mutations
110 (Hunter 2005). Although the GxE risk addition effects are reported to be potent and
111 widely existed, GxE effects are often hard to discover and therefore overlooked from
112 the scope of general pathophysiology. To this end, the advent of big data and the genetic
113 biobanking era has enabled diverse GxE studies in the development of complex disease
114 (McAllister et al. 2017). While environmental stimuli such as ultraviolet radiation, air
115 pollutants (such as PM2.5, CO, and sulfates), and environmental toxins have been
116 extensively studied in the context of GxE mechanisms (Virolainen et al. 2022), scant
117 attention has been paid to the environmental effect of hot spring exposure—which
118 allegedly can bring forth health benefits, hence we set goal to understand the hot-spring
119 related GxE interactions.

120 Throughout ancient cultures ranging across the empire of Rome, China, Japan, and
121 the islanders of Austronesian, hot springs bathing was deemed to improve health, gain
122 cosmetic benefit, and expand life. Recent scientific studies have gathered
123 complementary evidence indicating that hot spring bathing could prevent diseases,
124 alleviate musculoskeletal discomfort, modulate nerve and endocrine systems, and
125 promote immunity against chronic diseases (such as whip syndrome, rheumatoid

126 arthritis (RA), cancer, hemorrhoids, and skin diseases) (Bidonde et al. 2014;
127 Verhagen et al. 2015). Moreover, Onsen therapy was certified by the Japanese Society
128 for Complementary and Alternative Medicine (JCAM) (Serbulea and
129 Payyappallimana 2012), and was subsidized by health insurance in European
130 countries (such as France, Germany, and Italy).

131 Hot spring is a geothermal heated water body that exerts indirect health influences
132 through remodeling the local atmosphere and local hydrology. Bioactive gaseous
133 substances such as H₂S generally range between 1.0 and 3.0 µg/m³ in urban areas,
134 whereas the H₂S level spiked as high as 1500 µg/m³ in the residential vicinity of
135 geothermal (Gorini et al. 2020). The exact effect of these bioactive geothermal-released
136 substances was not fully understood, but independent pilot studies had reported
137 positive association between H₂S exposure and hospitalization events related to
138 circulatory systems (Standardized Hospitalization Ratios (SHR)=1.04; 95%
139 CI=1.01–1.07) (Nuvolone et al. 2019). Consistently, another geothermal epidemiology
140 study in Italy also showed that the increased H₂S exposure in men could result in higher
141 cardiovascular mortality rates (odds ratio (OR)=1.22, 90% CI=1.03-1.44) (Nuvolone
142 et al. 2020). Furthermore, a cohort study in Rotorua (New Zealand) showed that dwelling
143 in geothermal emission areas was associated with increased malignancy rate and a 1.26
144 (95% CI=1.14–1.38) standardized incidence ratio (SIR) for cataract. (Bates et al. 1998)

145 While hot spring bathing was reported to bring folk-healing benefits to musculoskeletal
146 diseases, hot spring residency was also linked to cardiovascular threats, therefore the
147 health influence of hot spring exposure remains controversial to date.

148 In this study, we set goal to investigate the crossover effect between hot spring
149 residency and genetic predisposition. To be specific, we performed a three-component
150 study that included environmental association discovery, variant-disease association
151 discovery, and GxE interaction analysis to delineate the environmental effects of hot
152 spring exposure on orthopedic, ophthalmic, and cardiopulmonary diseases.

153

154 **2. Materials and Methods**

155 **2.1. Study Design**

156 This is a three-component study that includes environmental association discovery
157 (demographic analysis and logistic regression model), variant-disease association
158 discovery (exposure stratified GWAS, protein-protein interaction-network analysis
159 and polygenic risk score), and the GxE interaction analysis (crossover GxE effects
160 and multivariate accumulative analysis).

161

162 **2.2. Study Population**

163 Nation-wide TWB1 and TWB2 (<https://www.twbiobank.org.tw/>) participant registry

164 was obtained. The 129451 individuals were enrolled from the general population by
165 TWB1 and TWB2 between 2016 to 2023. At the time of entering the cohort, all
166 participants were cancer-free and were obligated to complete a structured questionnaire
167 that documented: sociodemographic characteristics, lifestyle, dietary habits,
168 environmental exposures, and major medical histories (Fanet al. 2008). The disease
169 case labels in TWB were self-reported, and control cases were designated by
170 disease-free individuals that were above 60 years old.

171

172 **2.3.Outcomes of Interest**

173 The primary outcomes of interest of this study was set on hot-spring related orthopedic
174 diseases (osteoporosis, arthritis, gout), ophthalmic diseases (cataracts, glaucoma, dry
175 eye disease; DED, retinal detachment; RD, floaters, blindness, color vision deficiency;
176 CVD and age-related macular degeneration; AMD), cardiopulmonary diseases (asthma,
177 valvular heart disease; VHD, coronary artery disease; CAD, arrhythmia,
178 cardiomyopathy, congenital heart disease, hyperlipidemia, hypertension, stroke,
179 diabetes).

180

181 **2.4.Exposure of Interest**

182 The exposure of interest in this study includes hot spring area residency and

183 polygenetic underlying with adjustment of socioeconomic confounder.

184 For which we divided hot-spring and non-hot-spring residential areas by the

185 proclamation of the National water resource department. The hot spring definitions were:

186 water body heated above 30 celsius with total dissolved solid >500 mg/L and meeting

187 one additional minor criterions in category (I-a) $\text{HCO}_3^- > 250 \text{ mg/L}$ (I-b) $\text{SO}_4^{2-} > 250$

188 mg/L (I-c) Cl^- and other halides $> 250 \text{ mg/L}$, or (II-a) free $\text{CO}_2 > 250 \text{ mg/L}$ (II-b) total

189 sulfide $> 0.1 \text{ mg/L}$ (II-c) total Fe^{2+} and total $\text{Fe}^{3+} > 10 \text{ mg/L}$ (II-d) Radium $> 10^{-8} \text{ curie/L}$.

190 The polygenetic underlying was primarily assessed through variant-disease association

191 (VDA), which identified disease-associated SNPs through GWAS. Secondly,

192 independent polygenic risk score (PRS) models were established based upon the

193 GWAS summarized SNP lists, whereon each PRS predicts the corresponding disease

194 risk for hot spring residents and non-hot spring residents. The functional PRS SNP

195 genes were further analyzed with gene ontology (GO), co-expressed signaling

196 pathways, and protein-protein interaction (PPI) network analysis.

197

198 **2.5.Adjusted Covariates for Hot spring Exposure and Disease Outcome**

199 We first performed descriptive analysis. Continuous data are presented as means with

200 standard deviation, and categorical data are presented as proportions. Our study used

201 the Student's t-tests (normally distributed data) or Wilcoxon tests (non-normally

202 distributed data) for continuous variables and chi-squared tests or fisher exact test for
203 categorical variables between two groups (hot spring residents vs controls).

204 To find hot spring exposure related diseases, odds ratio (OR) and logistic regression
205 were conducted to detect disease association with hot spring residency, a subsequent
206 multivariate analysis was performed to solidify disease association under the
207 consideration of covariate adjustments. Since hot springs are mainly distributed in
208 townships with deviated sex ratios, imbalanced age composition, sparse medical
209 institutes, and less income, these socioeconomic and demographic covariates may
210 confound with the relationship between hot spring exposure (E) and the disease
211 outcome (D) (Supplementary Table 1 and Figure 1).

212 To specify the data source of the adjusted covariates: the demographic covariates,
213 including age, sex ratio and total resident counts, were recorded from the
214 Directorate-General of Budget, Accounting and Statistics. Total comprehensive income
215 was collected from the released tolls of the Ministry of Finance
216 (<https://www.mof.gov.tw/>). Accessible medical resources were estimated by health
217 insurance index (sum of health insurance benefits spent among all the hospitals in the
218 township) and medical visit index (sum of annual medical visits among all the hospitals
219 in the township) from the auditory releases of the Ministry of Health and Welfare
220 (<https://data.gov.tw/dataset/39280>). Continuous variables that did not follow a normal

221 distribution were provided with the mean±standard deviation (SD) and quartile-based
222 (min, Q1, Q2, Q3, and max) categorical conversion. The first category was pinned as
223 reference, then the effect of covariates was evaluated by univariable and multivariable
224 logistic regression models.

225 Univariate logistic regression analyses were used to show the significant and
226 independent role of the different variables in determining outcome (including arthritis,
227 floaters, DED, and VHD) effects. Multivariate logistic regression analyses were used to
228 control for confounding variables. All variables were then used for a multivariate
229 logistic regression model with stepwise selection for constructing prediction models.

230 Since in logistic regressions, estimated coefficients cannot be interpreted as a measure
231 of the contribution of the effect, we have also calculated marginal effects. The marginal
232 effect reflects the association between a small change in a variable and the alteration in
233 probabilities across each outcome. These marginal effects were computed while
234 keeping all other variables constant at the mean of the entire sample. Subsequently,
235 multivariable logistic regression analyses were employed to investigate potential
236 gene-environment interactions, utilizing the most suitable genetic model available. The
237 predictive performance of the constructed predictive models was evaluated using
238 sensitivity, specificity, and areas under the receiver operating characteristic (ROC)
239 curves, the AUC value. All statistical analyses were performed using R version 4.1.1.

240 The tests were 2-tailed, and p -value < 0.05 was taken as statistically significant.

241

242 **2.6. Variant-disease Association Discovery**

243 All SNPs discoveries were performed on PLINK (version 1.9 downloaded from

244 <http://zzz.bwh.harvard.edu/plink/>). SNPs quality control was conducted with the

245 following criteria: (i) SNPs missing in more than 2% of participants (ii) individuals

246 missing more than 5% SNP data (iii) SNPs with minor allele frequency (MAF) under

247 0.05 (iv) SNPs deviating from Hardy–Weinberg equilibrium (p -value < 0.05) (v)

248 individuals with identity by descent (IBD) over 0.125.

249 GWAS identified disease-specific SNPs for populations living in townships with and

250 without hot springs. SNPs were further screened by pairwise squared correlation (r^2)

251 less than 0.01 that linkage disequilibrium (LD) was estimated with a window size of

252 4000 bp. Functional SNP genes were labeled as G_A for significant SNPs from all

253 populations, G_N for significant SNPs from non-hot spring populations; G_H for

254 significant SNPs from hot spring populations, and G_h indicated hot spring-specific

255 functional SNP genes G_H that did not overlap in G_A and G_N . Furthermore, based upon

256 the GWAS summarized SNP lists, each PRS predicts the corresponding disease risk for

257 hot spring residents and non-hot spring residents.

258 The functional PRS SNP genes were further analyzed on Enrichr (Chenet al. 2013;

259 Kuleshov et al. 2016; Xie et al. 2021) to attribute gene ontology (GO) and the ARCHS4
260 Kinases Coexp library was applied to identify co-expressed signaling pathways.
261 Protein-protein interaction (PPI) network analysis of G_h was further evaluated in the
262 Search Tool for Retrieval of Interacting Genes/Proteins (STRING version 11.5
263 download from <http://string-db.org/>) (Szklarczyk et al. 2018) with a confidence level of
264 0.400 (medium) based on the find of: experiments, databases, co-expression,
265 neighborhood, gene function, and co-occurrence. Then, the activated pathways in the
266 selected PPI network were further calculated with the Reactome 2022 database.

267

268 **2.7. GxE Interaction Analysis**

269 Crossover GxE effects between environmental conditions and individual genetic
270 variants were assessed by relative risk (RR) with hot spring and PRS score as binary
271 variables: living in township with/without hot spring and high/low PRS score.
272 Visualized by nomogram, multivariate accumulative analysis based on logistic
273 regression model estimated predictability of confounders (sex) and GxE effect.
274 Decision curve analysis (DCA) was performed to evaluate a clinical “net benefit” for
275 nomogram prediction models in comparison to default strategies of treating all or no
276 patients. Net benefit is calculated across a range of threshold probabilities, defined as
277 the minimum disease risk at which further intervention would be warranted.

278 Net benefit = sensitivity × prevalence – (1 – specificity) × (1 – prevalence) × w

279 where w is the odds at the threshold probability.(Vickers et al. 2019)

280

281 **3. Results**

282 **3.1.Finding Hot Spring Associated Disease**

283 This study included 129,451 participants with 4517 individuals dwelled in the

284 townships with hot spring outcrops; their demographic characteristics were detailed as

285 in Table 1.

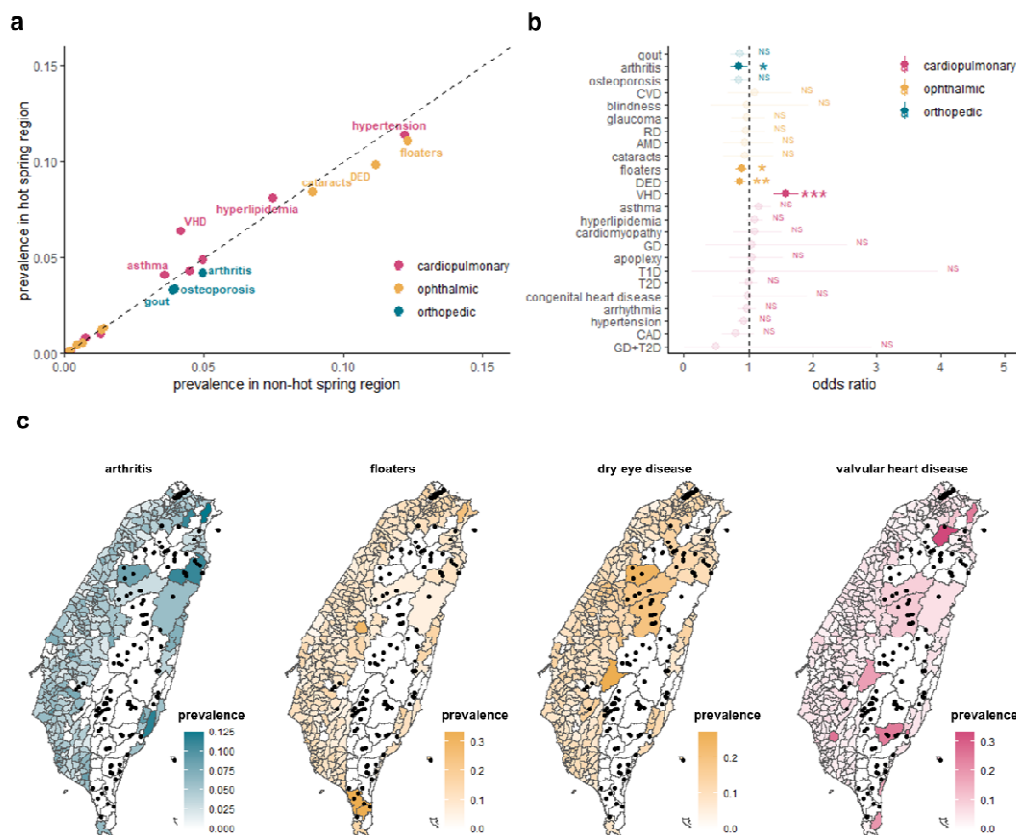
286

287 **Table 1: Demographic Information of the Participants.**

	Hot spring residents (n=4517)	Control (n=124934)	Number (%)	p-value
Age (years)	48.56±10.94	49.76±10.96	129451 (100.0)	<0.001
Body mass index(kg/m²)	24.03±3.40	24.00±3.40	126783 (97.9)	0.682
Water daily intake (ml)	1286.80±633.15	1255.12±685.91	21735 (16.8)	0.07
Sex				
Male (%)	1658 (36.7)	45066 (36.1)	129451 (100.0)	0.392
Female (%)	2859 (63.3)	79868 (63.9)		
Alcoholism				
No or seldom (%)	4122 (91.3)	114134 (91.4)	129338 (99.9)	0.578
Abstain (%)	111 (2.5)	3268 (2.6)		
Yes (%)	282 (6.2)	7421 (5.9)		
Smoking experience				
Yes (%)	3213 (71.2)	90795 (72.7)	129404 (100.0)	0.025
No (%)	1301 (28.8)	34095 (27.3)		
Water origin				
Tap water (%)	77 (1.7)	2716 (2.2)	21958 (17.0)	0.018
Shallow well water (%)	0 (0.0)	13 (0.0)		
Deep well water (%)	9 (0.2)	126 (0.1)		
Mineral water (%)	35 (0.8)	1125 (0.9)		
Filtered water (%)	390 (8.6)	15934 (12.8)		
Other (%)	49 (1.1)	1484 (1.2)		

288 Among the included disease spectrum, four diseases (arthritis, floaters, dry eye
 289 disease and valvular heart disease) showed higher prevalence in hot-spring areas.
 290 (Figure 1a and 1b). However, the coefficient of determination of these diseases were

291 not significantly associated with hot spring distribution (Figure 1c; $R^2=4.40\%$ in
 292 arthritis, 2.51% in floaters, 1.54% in DED, and 4.53% in VHD)



293

294 **Figure 1: Demographic Analysis Finding Environmental (E) Effect.**

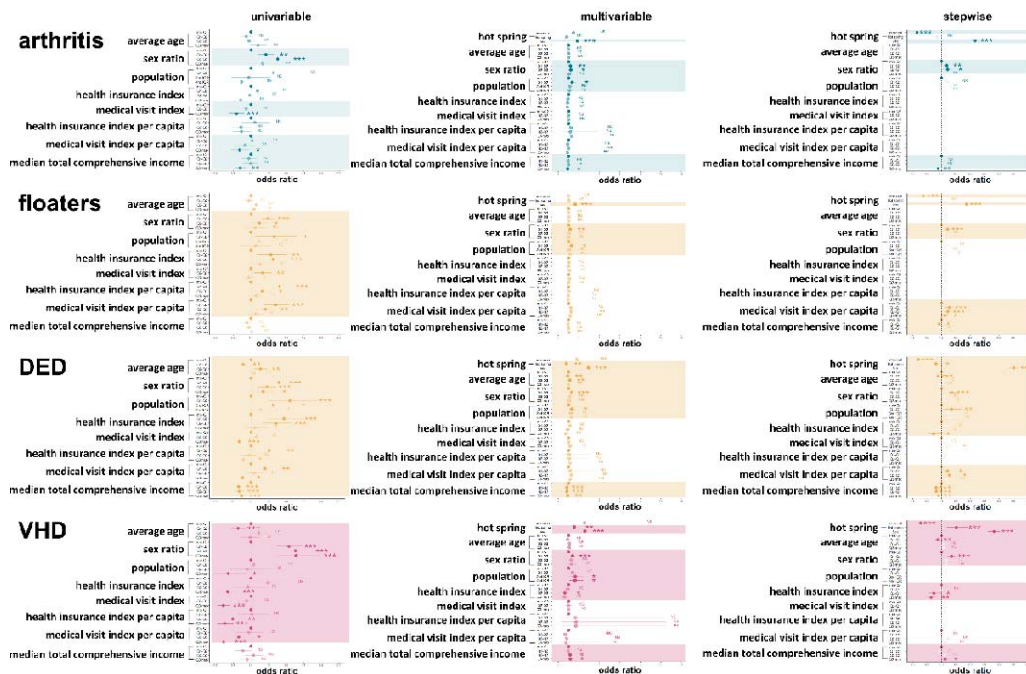
295 (a) Comparing the disease prevalence between hot spring region and non-hot spring region. The dashed
 296 line indicated the same disease prevalence in both regions. (b) Odds ratio (OR) indicating associations
 297 between hot spring geology and disease prevalence with 95% confidence interval (CI). The dashed line
 298 indicated OR=1.0 (c) Prevalence distribution of hot spring-related disease. The dots indicated the
 299 location of the hot spring outcrop. DED, dry eye disease; RD, retinal detachment; CVD, color vision
 300 deficiency; AMD, age-related macular degeneration; VHD, valvular heart disease; CAD, coronary
 301 artery disease; GD, gestational diabetes (GD); T1D, type 1 diabetes; and T2D, type 2 diabetes.

302 After univariate and multivariate logistic regression, DED (multivariable: OR=0.78,

303 95% CI=0.65-0.92; stepwise: OR=0.81, 95% CI=0.679-0.961) and VHD

304 (multivariable: OR=1.56, 95% CI=1.20-2.03; stepwise: OR=1.52, 95% CI=1.20-1.91)

305 stood out as the two remaining hot spring-associated diseases (p -value<0.05); whereas
306 the hot spring association with arthritis (p -value=0.0589 for multivariable;
307 p -value=0.1496 for stepwise) and floaters (p -value=0.6807 for multivariable;
308 p -value=0.9935 for stepwise) were not statistically significant after adjusting for
309 social environmental factors (Figure 2). In these multivariate regression models, the
310 average marginal effects (AMEs) of hot spring were statistically significant for both
311 DED (AME=-0.23, 95% CI=-0.40 -0.07, p -value=0.0060) and VHD (AME=0.41,
312 95% CI=0.17-0.64, p -value=0.0008). Moreover, the incorporation of hot spring
313 exposure information enhanced the stepwise logistic regression model performance in
314 DED and VHD prediction (Table S2). These statistic results indicated a robust
315 association between disease (DED and VHD) risk modification and hot spring
316 distribution.



317

318 **Figure 2: Logistic Regression Models Evaluating Socioeconomic (S) Confounding Effects.**

319 The dashed line indicated odds ratio (OR)=1.0 DED, dry eye disease; and VHD, valvular heart disease.

320

321 3.2. Variant-disease Association Discovery

322 We applied GWAS (SNP list: Supplementary Table 3-8) and Manhattan plot to

323 visualize the genetic predisposition of the DED and VHD. Interestingly, from the

324 designated subgroup analysis, we noted that the significant (p -value $<10^{-5}$) SNPs from

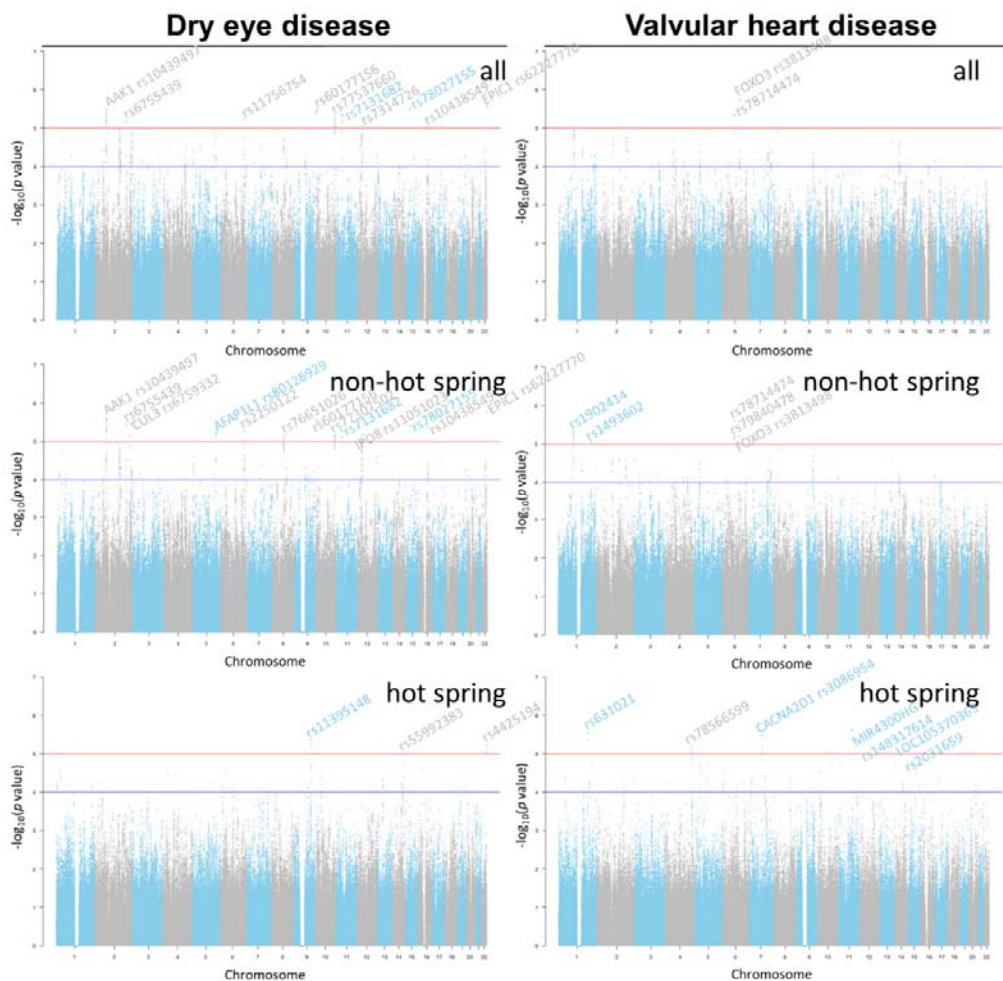
325 non-hot spring subgroup were partial overlay with the parental all participants (for

326 instance, DED: AAK1 rs6755439, rs60177156, rs7131682, rs78027155, rs10438549,

327 *EPIC1* rs62227770, and so on; VHD: rs78714474 and *FOXO3* rs3813498) but were

328 distinct to the hot spring subgroup findings. This SNP discovery discrepancy indicated

329 a potential GxE effect (Figure 3).



330

331 **Figure 3: Manhattan Plots of Genome-wide Association Studies (GWASs) among “All”,**
 332 **“Non-hot Spring”, and “Hot Spring” Gene Sets.**

333 Blue lines represented p -value = 10^{-4} while red lines represented p -value = 10^{-5} .

334 We then calculated the predict AUC with the derived SNPs list at different significance

335 (Figure 4a). At p -value < 0.005, the predict AUC for DED hot spring and VHD hot

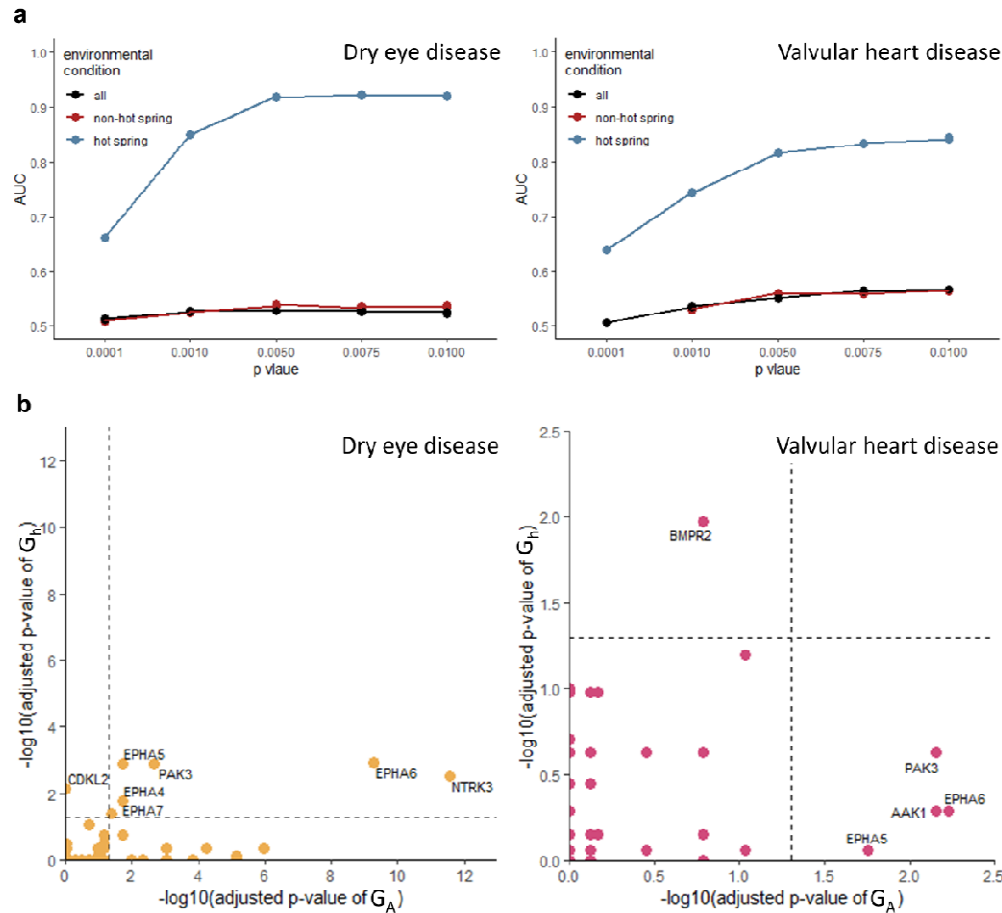
336 spring subgroup were 0.92 and 0.82 respectively. However, in the non-hot spring

337 subgroup, the model does not gain AUC increment over the enrolled SNPs.

338 After adjusting LD threshold ($r^2 < 0.01$) and genome-wide significance threshold

339 (p -value < 10^{-5}) (Supplementary Figure 2), the VDA results were aggregated and

340 annotated as G_A , G_N , and G_H for “all”, “non-hot spring”, and “hot spring” subgroup
341 respectively. 73 functional DED SNPs and 19 functional VHD SNPs among G_A , G_N ,
342 and G_H were overlapped (Supplementary Figure 3). After pathway discovery, there
343 were 15 and 4 significant ARCHS4 human kinase pathways in the G_A of DED and
344 VHD respectively (adjusted p -value<0.05) (Figure 4b). However, CDKL2 kinase
345 pathway was the only significant hit for G_h subgroup (Overlapped ratio: 21/299 with
346 *GABARAPL2*, *C11ORF87*, *CXADR*, *DOCK3*, *ANKRD46*, *DDX25*, *PTPRO*, *ANK2*,
347 *NALCN*, *LRP1B*, *MYO16*, *TRIM9*, *PCLO*, *CACNB4*, *DLG2*, *TRIM2*, *PSD3*, *TP53BP1*,
348 *CTNNA2*, *SRGAP3*, *CDH18*; OR=2.79, adjusted p -value=0.00698). Meanwhile,
349 BMPR2 kinase pathway was the only significantly hit in the VHD G_h (Overlapped ratio:
350 17/299 with *RBFOX1*, *EPAS1*, *CACNA2D1*, *CRIM1*, *PTPRM*, *PARVA*, *ZDHHC21*,
351 *LAMB1*, *ATP2B2*, *KALRN*, *HIPK2*, *SIPA1L1*, *AKT3*, *EDIL3*, *CLASPI*, *KCNH1*,
352 *OPCML*; OR=3.51, adjusted p -value=0.01066).
353



354

355 **Figure 4: Signaling Pathways of Co-expressed Genes based ARCHS4 Kinases Coexp Library.**

356 (a) Screening single nucleotide polymorphisms (SNPs) based on genome-wide significance level
 357 threshold ($p\text{-value} < 0.005$). AUC, Area Under Curve. (b) Comparing signaling pathways of
 358 co-expressed genes between gene set from all population (G_A) and hot spring-specific gene set (G_h). The
 359 dashed line indicated $p\text{-value} < 0.05$.

360 To investigate the putative molecular interaction in the derived G_h hits, we conducted

361 Reactome pathway detection in the PPI after filtering and clustering G_h based on

362 protein crosstalk in DED and VHD respectively. Consisting of 158 genes with *RYR2*,

363 *CACNA1C*, *TGF β 1*, and *MAPK8* as four central hub genes, activated pathways in PPI

364 network of DED were related to proposed DED-related pathomechanisms, such as

365 O-linked glycosylation of mucins (R-HSA-913709), Interleukin (IL)-4 and IL-13

366 signaling (R-HSA-6785807), and interaction between L1 and ankyrins
367 (R-HSA-445095)(Figure 5a and 5b). Consisting of 39 genes with *PARP1*, *GRM1*,
368 *NRXN2*, and *GRIA4* as four central hub genes, activated pathways in one of PPI
369 network of VHD were related to proposed VHD-related pathomechanisms, such as
370 Calcium (Ca^{2+}) channel opening (R-HSA-112308), YAP1- And WWTR1
371 (TAZ)-stimulated gene expression (R-HSA-2032785), thrombin signaling
372 (R-HSA-456926), cardiac conduction (R-HSA-5576891),Runt-related transcription
373 factor 2 (Runx-2) (R-HSA-8939243), mitogen-activated protein kinase (MAPK)
374 pathway (R-HSA-5674135), interaction between RAF1 and BRAF (R-HSA-5683057),
375 and signaling by GPCR (R-HSA-112308)(Figure 5c-1 and 5d-1). Comprised 6 genes
376 (*CNOT7*, *CPEB3*, *CNOT2*, *UBXN7*, *EPAS1*, and *COMMD10*), activated pathways in
377 another of PPI network of VHD were related to proposed VHD-related
378 pathomechanisms, such as autophagy (R-HSA-9612973) and hypoxia
379 (R-HSA-1234174) (Figure 5c-2 and 5d-2).



380

381 **Figure 5: Activated Pathways of Functional Genes within Protein-protein Interaction (PPI)**
 382 **Network of Hot Spring-specific Gene Set (G_H) based on Reactome 2022.**

383 (a) PPI of dry eye disease (DED). (b) Activated pathways of DED. (c) PPI of valvular heart disease
 384 (VHD). (d) Activated pathways of VHD. PARS, proteinase activated receptors; HIF1 α ,
 385 hypoxia-inducible factor-1 α ; MAPK, mitogen-activated protein kinase; and R-HAS, Reactome Homo

386 sapiens.

387 PRS models were constructed for different environmental conditions (all, non-hot

388 spring, hot spring) based on the associated SNPs discovered from the TWB (G_A , G_N ,

389 and G_H). The mean PRS was higher among the cases compared to the controls across

390 all models in both diseases but more significant in G_H (AUC: 0.9168 for DED and

391 0.8153 for VHD) (Figure 6a). With $MAF > 0.05$ and p -value of Hardy–Weinberg

392 equilibrium > 0.05 , SNPs used in PRS model were not rare variants and were

393 independent of environmental conditions (E). However, PRS models with G_H could

394 not predict disease risk well in populations without hot spring exposed (AUC: 0.5015

395 in DED and 0.5066 in VHD), indicating significant GxE relationship. The GxE

396 relationship demonstrated a dose-response effect (Figure 6b). Compared to the

397 participants with PRS in min-Q1, participants with PRS in Q2-Q3 demonstrated 29.11

398 times odds of developing DED (95% CI=10.36-81.76, p -value <0.001) while the

399 individuals with PRS in Q3-max demonstrated 390.81 times odds (95%

400 CI=130.93-1166.52, p -value <0.001) (Figure 6c). Compared to the participants with

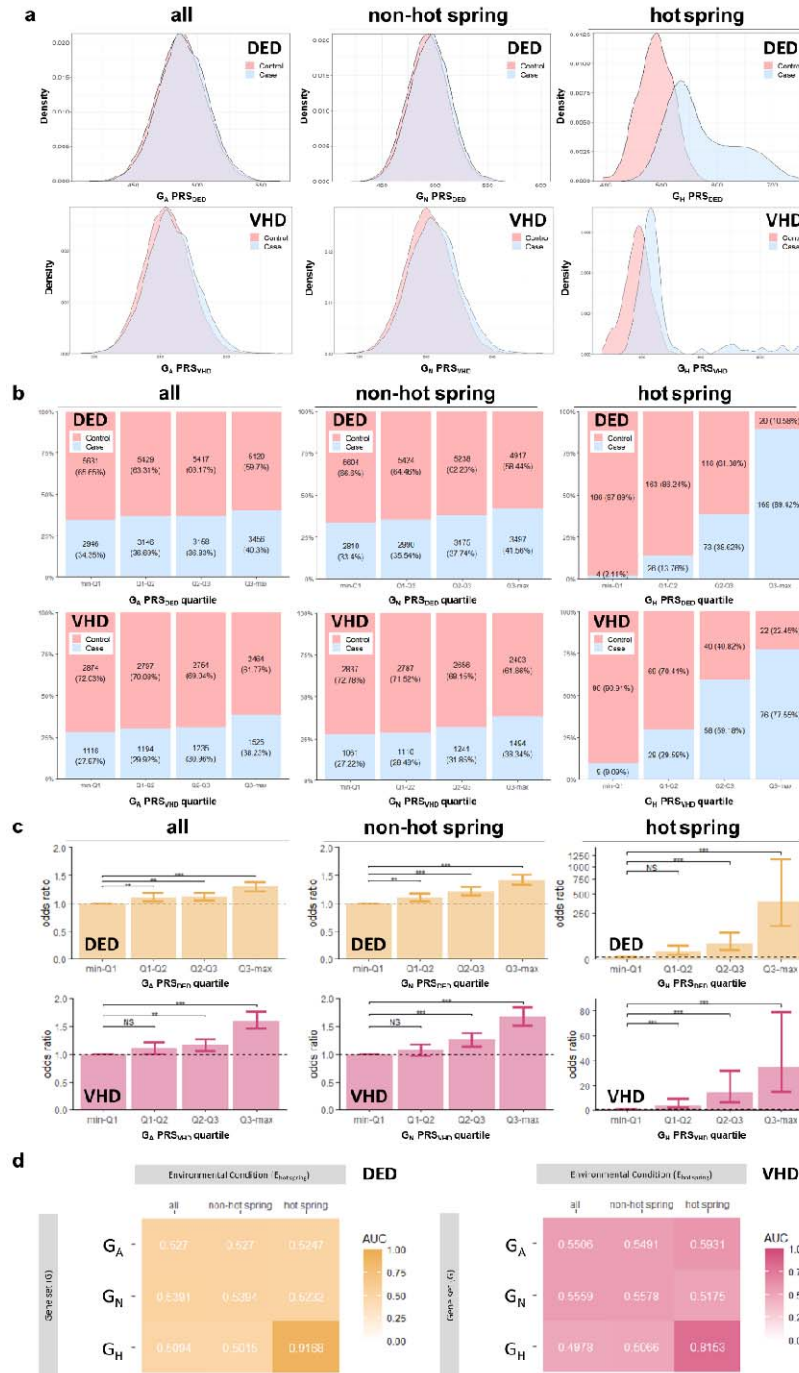
401 PRS in min-Q1, participants with PRS in Q2-Q3 demonstrated 14.339 times odds of

402 developing VHD (95% CI=6.475-31.754, p -value <0.001) while the individuals with

403 PRS in Q3-max demonstrated 34.162 times odds (95% CI=14.842-78.63,

404 p -value <0.001) (Figure 6c).

405



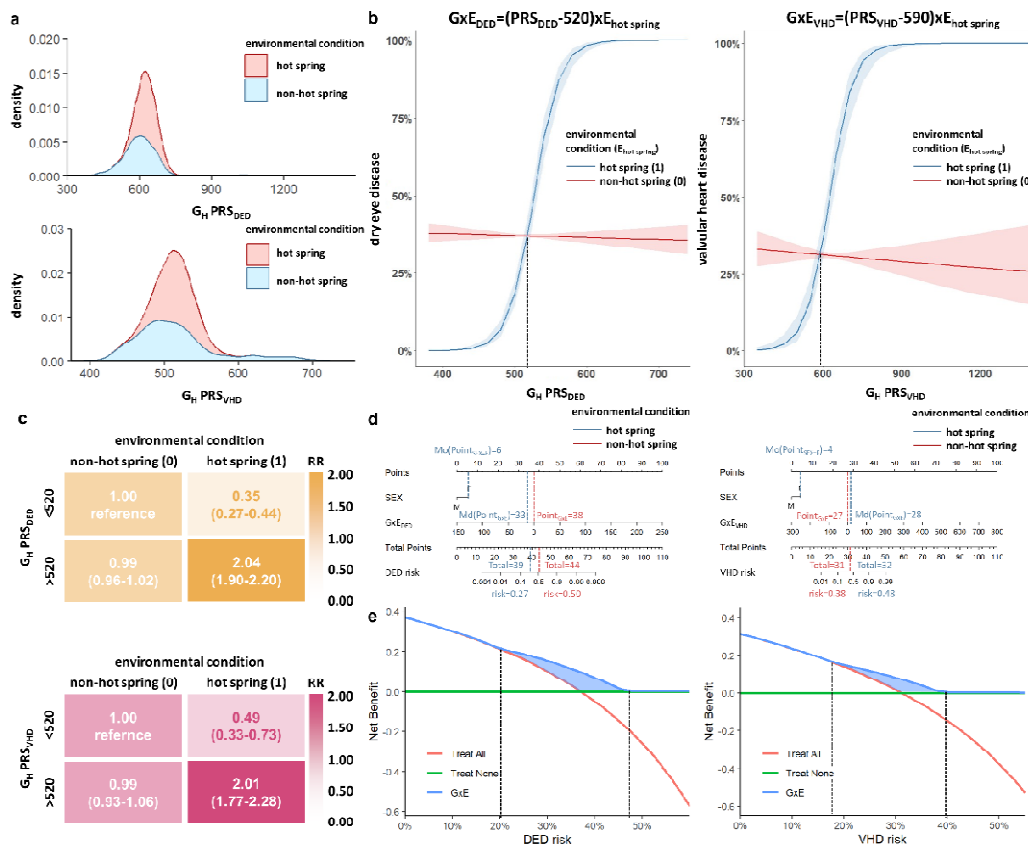
406

407 **Figure 6: Polygenic Risk Score (PRS) Models Evaluating Gene-environment interactions (GxE).**

408 (a) Density distribution among PRS score. After adjusted for linkage disequilibrium (LD) clumping
 409 threshold ($r^2 < 0.01$) and genome-wide significance level threshold ($p\text{-value} < 0.005$), variant-disease
 410 association (VDA) results were depicted as G_A , G_N , and G_H for “all”, “non hot spring”, and “hot spring”
 411 gene sets. (b) Percentage of case and control in each PRS quartile. (c) Odds ratio (OR) of each PRS
 412 quartile. The dashed line indicated OR=1.0 (d) Area Under Curve (AUC) of different combinations of
 413 gene set (G) and environmental condition (E). DED, dry eye disease; and VHD, valvular heart disease.

414 **3.3.GxE Interaction Analysis**

415 While participants with different environmental conditions showed similar density
416 profiles of G_HPRS (Figure 7a), the disease ratio was significantly different with
417 significant GxE interaction indicated by logistic regression model (p -value $<10^{-16}$),
418 showing crossover GxE effect (Figure 7b). Since disease ratio almost remained
419 unchanged across G_HPRS, GxE effect was calculated based on risk in non-hot spring
420 region as reference. RR tables between different environmental conditions and
421 G_HPRS indicated hot spring as effect modifier (Figure 7c). Therefore, we enrolled
422 gender and GxE effect to construct DED and VHD predict nomograms (Figure 7d).
423 The DCA curve was performed to demonstrate high clinical net benefit between
424 threshold probability of 20%-48% for DED and 18%-39% for VHD (Figure 7e).
425



426

427

Figure 7: Net Benefit of Identifying the Risk of Gene-environment Interactions (GxE).

428

(a) Density profile of G_H polygenic risk score (PRS). (b) Multivariate logistic regression evaluating the

429

GxE effect. (c) Relative risk (RR) with 95% confidence interval (CI) between different environmental

430

conditions and G_H PRS. (d) Nomogram results estimating predictability of GxE. (e) Decision curve

431

analysis (DCA) assessing net benefit of identifying GxE risk. PRS, polygenic risk score; Mo, mode;

432

Md, median; G_H , gene set from “hot spring”; DED, dry eye disease; and VHD, valvular heart disease.

433

434 4. Discussion

435

By employing multimodal analysis based on genetic data in Taiwan biobank and DA in

436

Opendata platform, we conducted environmental association discovery, variant-disease

437

association discovery, and the GxE interaction analysis to reconstitute GxE effect on

438

DED and VHD. This context dependent GxE interactions were discussed as below:

439

440 **4.1.Hot Spring and Disease Correlation**

441 In the environmental association discovery, residency in hot spring areas was
442 associated with a lower prevalence of dry eye (OR=0.81) and a higher prevalence of
443 valvular heart disease (OR=1.52).

444

445 *4.1.1. Hot Spring-related Factors in DED*

446 Dry eye is a result of increased osmolarity of the tear film and inflammation of the
447 ocular surface (Pflugfelder and de Paiva 2017). It can be influenced by external factors
448 such as desiccating environment and wind exposure (Tsubota et al. 2017). In previous
449 air pollutants studies, higher humidity levels were associated with improved corneal
450 fluorescein staining and tear breakup time (TBUT) (Berget et al. 2020), while participants
451 residing in areas with increased relative humidity were less likely to develop DED
452 (Hwanget et al. 2016). Given that hot spring water vapor may increase humidity levels in
453 the immediate vicinity (Nordstrom et al. 2005), it is possible that residing near hot
454 springs were less likely to develop dry eyes. To the best of our knowledge, there are no
455 studies publicly available to study the atmospheric CO₂ H₂S and N₂O influence on
456 DED, future research avenue in the atmospheric (SO₂, CO₂ H₂S, N₂O) effect on DED
457 occurrence.

458

459 *4.1.2. Hot Spring-related Factors in VHD*

460 Regarding VHD, research has demonstrated that environmental factors, specifically
461 Legionellosis outbreaks in regions such as Taiwan (Hsuet al. 2006) and Japan (ITO et al.
462 2002), have been linked to the transmission of *Legionella* through hot springs. The
463 Legionella bacteremia often result from pulmonary sources and may cause endocarditis
464 and valvular infections (Brusch 2017). Additionally, other microorganisms, such as
465 *Acanthamoeba spp.* and *Microspora*, have also been associated with hot spring-related
466 transmission and could potentially lead to cardiac complications (Filho et al. 2009). The
467 positive correlation of hot spring distribution and valvular heart disease may be due to
468 ambient pollutants emitted from water or infectious disease spreading through the
469 hydrosphere.

470

471 **4.2. Genetic Effects in DED and VHD**

472 In variant-disease association discovery, we demonstrated that co-expression of kinase
473 pathways (CDKL2 for DED and BMPR2 for VHD) and disease-related reactome
474 pathways were significantly enriched in hot-spring specific functional SNPs.

475

476 *4.2.1. Hot Spring-specific SNP Genes Co-expressed Kinase Pathway–CDKL2 in DED*

477 Cyclin-dependent kinase-like 2 (CDKL2) belongs to the family of CDC2-related

478 serine/threonine protein kinases. CDKL2 was identified as a target related to Sjögren's
479 syndrome (Xia et al. 2023) and rheumatoid arthritis (RA) was the most enriched term in
480 the KEGG analysis of CDKL2 methylation geneset (Chenet et al. 2021). These studies
481 supported the association between CDKL2 and DED, as dry eye is the most frequent
482 ocular manifestation in patients with Sjögren's syndrome (>90%) and RA (46.7%)
483 (Kosrirukvongset et al. 2012). The co-expression of CDKL2/CDKL3 (Tracey et al. 2003)
484 and interleukins (IL-15RA, IL-7, IL-1RAPL1, and IL-11) also suggested correlation
485 between CDKL2 and DED as Sjögren's syndrome, RA, and DED are known to be
486 mediated by cytokine-mediated systemic inflammation. Additionally, CDKL2
487 expression was detected in DED-involved ocular segments such as the lacrimal gland,
488 cornea, and eyelid, according to the Human Eye Transcriptome Atlas (Tracey et al.
489 2003). Although a direct link between CDKL2 mutation and DED was not found in
490 current research, we postulate that the effect of CDKL2 on DED is context-dependent
491 and influenced by additional environmental cues.

492

493 *4.2.2. Activated Pathway in G_h SNPs of DED*

494 Activated pathways in the PPI network of DED included O-linked glycosylation of
495 mucins, IL-4 and IL-13 signaling, and interaction between L1 and ankyrins.
496 Transmembrane mucins (such as MUC1, MUC4, and MUC16) (Uchino 2018), large

497 glycoproteins with heavily glycosylated glycans, were revealed to be essential for
498 maintaining ocular surface epithelium lubrication, wettability, and epithelial
499 glycocalyx barrier. In response to inflammatory stimulation of DED, goblet cells and
500 corneal/conjunctival epithelial cells express receptors for inflammatory cytokines such
501 as IL-13 (Baudouinet al. 2019), inducing an important transcriptional factor in
502 epithelial differentiation and goblet cell formation. Activation of transient receptor
503 potential ankyrin 1 (TRPA1) resulted in calcium influx, tear reduction, increased eye
504 blink, and forelimb eye wipe behavior.(Ashoket al. 2023) These G_n-related pathways
505 implied the underlying pathomechanisms of GxE in DED.

506

507 *4.2.3. Hot Spring-specific SNP Genes Co-expressed Kinase Pathway–BMP2 in VHD*

508 Bone morphogenic receptor 2 (*BMP2*) gene encode a transmembrane
509 serine-threonine kinase receptor that governs SMAD signaling pathways. While
510 SMAD signal cascade was associated with aortic dilation (Balintet al. 2022),
511 dysregulation of the upstream BMP2 kinase pathways may hold valuable study
512 directions in GxE VHDs. BMP2 loss of function SNPs were discovered in heritable
513 pulmonary arterial hypertension (PAH) and idiopathic PAH (Chenet al. 2019), these
514 two PAH conditions concurs in 15-60% VHD patients (Magneet al. 2015). Li et al. was
515 the first to report BMP2 may cause VHD by DNA hypermethylation through

516 down-regulation of *BMPR2* expression (Liet al. 2021). In echo to this report, our result
517 identified *BMPR2* pathway as a prominent functional SNP pathway in the hot
518 spring-VHD patients, whereas *BMPR2* pathway was not evident in general VHD
519 patients.

520

521 *4.2.4. Activated Pathway in G_h SNPs of VHD*

522 Activated pathways in the PPI network of VHD included Ca^{2+} channel opening, YAP1-
523 And WWTR1 (TAZ)-stimulated gene expression, Runx-2, and MAPK pathway.
524 Recently, combining 2 GWAS studies in 474 and 486 cases from Canada and France, a
525 meta-analysis confirmed the role of RUNX2 and CACNA1C as susceptibility genes of
526 aortic valve stenosis (AS), belonging to the Ca^{2+} signaling pathway (Guauque-Olarteet
527 al. 2015). Moreover, YAP1 has been reported to regulate endothelial to mesenchymal
528 transition (EMT) through modulation of TGF β -Smad signaling and proliferative
529 activity during cardiac cushion development. MAPK pathway influences calcification,
530 being linked to the expression of a contractile phenotype in valvular interstitial cells.
531 These G_h -related pathways implied the potential mechanisms of GxE in VHD.

532

533 **4.3.GxE Interaction and Clinical Prospect**

534 In this study, we reported that hot spring residency and underlying polygenetic

535 conjointly modified the disease risk of DED and VHD, whereupon we gathered
536 multi-angle evidence to strengthen the significance of GxE interaction. Although it is
537 generally accepted that both genetic and environmental factors contribute to the
538 development of complex diseases, few GxE examples were replicated, biologically
539 plausible, and methodologically sound interaction with proven clinical relevance and
540 application in daily clinical routines. Therefore, the term “GxE interaction” was often
541 used to express that several factors contribute to disease risk without excluding the
542 possibility of complete independence. To be specific, biological GxE interaction is the
543 joint effect of both genetic and environmental factors that act together in a direct
544 physical or chemical reaction or in the same causal mechanism of disease development
545 (Yang and Khoury 1997) while statistical GxE interaction is defined as “departure from
546 additivity of effects on a specific outcome scale (Rothman and Greenland 1998).” That
547 is, the effect of one factor depends on the level of the other factor. The interaction could
548 be defined as crossover effect or nonremovable interaction only when the interaction
549 remained in any monotone transformation (Thompson 1991). In this study, the
550 interaction of hot spring residency and the level of GH PRS well complied to the
551 requirements of statistical GxE interaction (Figure 7b). However, neither long-term
552 residency effects nor diseases-specific effects have been thoroughly investigated with
553 hot spring. Therefore, we struggled to interpret the biological GxE interaction via GO,

554 co-expressed signaling pathways, and PPI network analysis. Overall, we proposed a
555 novel GxE interaction in hot spring-associated diseases, whereby we suggest further
556 studies are warranted to elucidate the benefits and harms of long-term hot spring
557 residency in human health.

558

559 **4.4. Study Limitation**

560 Despite the favorable results, there are several study limitations in this study. First, the
561 acquisition of VHD and DED disease labels relied on self-reported questionnaires; the
562 definitive diagnosis records and degree of disease severity were intrinsically absent
563 from the biobank data. The hot spring exposure was determined by participant
564 residencies, the exact exposure duration and intensity cannot be estimated and studied
565 from existing accessible data. Although multi-angle evidence supported the GxE
566 interaction in hot spring-related DED and VHD, the concise pathomechanism cannot
567 be inferred from statistical and epidemiological studies. Together, this study served as
568 a pilot study that highlights potential research avenues for GxE effects in hot
569 spring-related diseases.

570

571 **5. Conclusion**

572 Our findings validate the pivotal role of GxE interactions in the epidemiology of DED

573 and VHD within Taiwan's hot spring regions, underscored by the enhanced disease
574 prediction through PRS models. The study emphasizes the complex interdependencies
575 between genetics and environmental exposures in disease manifestation, reinforcing
576 the need for integrative risk assessments in environmental health studies.

577 **Statements**

578 **Ethics approval and consent to participate**

579 Ethical approvals were obtained from the Institutional Review Board (IRB) of the
580 Taipei Veterans General hospital with reference numbers 2023-01-006AC. All
581 participants provided written informed consent. This study was conducted in
582 compliance with the Helsinki Declaration.

583

584 **Data Sharing**

585 All data is published in this manuscript, the cited manuscripts, or the supplementary
586 appendix. Data can be provided upon request to corresponding authors, and in
587 agreement of terms. The datasets generated and/or analyzed during the current study
588 are available in the Nation-wide Taiwan Biobank1 and Taiwan Biobank2
589 (<https://www.twbiobank.org.tw/>), Directorate-General of Budget, Accounting and
590 Statistics (<https://eng.dgbas.gov.tw>) Ministry of Finance (<https://www.mof.gov.tw/>),
591 Ministry of Health and Welfare (<https://data.gov.tw/dataset/39280>,
592 <https://data.gov.tw/dataset/39281>, <https://data.gov.tw/dataset/39282>,
593 <https://data.gov.tw/dataset/39283>) repository.

594

595 **Declaration of Interests**

596 The authors declare that they have no competing interest.

597

598 **Authors' Contributions**

599 HYW, KJC and YPY were responsible for the conceptualization and were actively
600 involved in the planning of methodology, investigation, and project administration.

601 PHC and HYT contributed to the formal analysis. HYW, KJC, WC, CYW participated
602 by validation, visualization, writing original draft, reviewing, and editing. YTH,

603 YCW, YCC, CHH, ARH, SHC and CCH provided critical advice. YHC, HJD, CHL,
604 and YCC accessed and verified the underlying data reported in the manuscript. All

605 authors had full access to all the data and responsibility for the decision to submit for
606 publication.

607

608 **Acknowledgements**

609 This study is based in part on data from the Big Data Center, Taipei Veterans General
610 Hospital (BDC, TPEVGH). The interpretations and conclusions contained herein do

611 not represent the position of Taipei Veterans General Hospital. The authors thank
612 Department of Statistics, Tamkang University for technological support.

613

614 This work was supported by the Taiwan National Science and Technology Council

615 [grant numbers NSTC 112-2321-B-A49-007, NSTC 111-2320-B-A49-028-MY3,
616 NSTC 112-2124-M-038-001, and NSTC 112-2314-B-032-001]; and Taipei Veterans
617 General Hospital [grant number V113C-201, V112C-026 and 112VACS-007].

618

619 **Reference**

620 Ashok, N.; Khamar, P.; D'Souza, S.; Gijs, M.; Ghosh, A.; Sethu, S.; Shetty, R. Ion
621 channels in dry eye disease. *Indian Journal of Ophthalmology* 2023;71:1215-1226.

622 https://doi.org/10.4103/IJO.IJO_3020_22

623 Balint, B.; Federspiel, J.; Kollmann, C.; Teping, P.; Schwab, T.; Schäfers, H.-J.; Balint, B.;

624 Federspiel, J.; Kollmann, C.; Teping, P.; Schwab, T.; Schäfers, H.-J. SMAD3 contributes
625 to ascending aortic dilatation independent of transforming growth factor-beta in

626 bicuspid and unicuspid aortic valve disease. *Scientific Reports* 2022;12:15476.

627 <https://doi.org/10.1038/s41598-022-19335-w>

628 Bates, M.N.; Garrett, N.; Graham, B.; Read, D. Cancer incidence, morbidity and
629 geothermal air pollution in Rotorua, New Zealand. *Int J Epidemiol* 1998;27:10-14.

630 <https://doi.org/10.1093/ije/27.1.10>

631 Baudouin, C.; Rolando, M.; Benitez Del Castillo, J.M.; Messmer, E.M.; Figueiredo, F.C.;

632 Irkec, M.; Van Setten, G.; Labetoulle, M. Reconsidering the central role of mucins in
633 dry eye and ocular surface diseases. *Prog Retin Eye Res* 2019;71:68-87.

- 634 <https://doi.org/10.1016/j.preteyeres.2018.11.007>
- 635 Berg, E.J.; Ying, G.S.; Maguire, M.G.; Sheffield, P.E.; Szczotka-Flynn, L.B.; Asbell, P.A.;
- 636 Shen, J.F.; Group, D.S.R. Climatic and Environmental Correlates of Dry Eye Disease
- 637 Severity: A Report From the Dry Eye Assessment and Management (DREAM) Study.
- 638 Transl Vis Sci Technol 2020;9:25. <https://doi.org/10.1167/tvst.9.5.25>
- 639 Bidonde, J.; Busch, A.J.; Webber, S.C.; Schachter, C.L.; Danyliw, A.; Overend, T.J.;
- 640 Richards, R.S.; Rader, T. Aquatic exercise training for fibromyalgia. Cochrane Database
- 641 Syst Rev 2014;2014:CD011336. <https://doi.org/10.1002/14651858.CD011336>
- 642 Bruschi, J.L. Legionnaire's Disease. Infectious Disease Clinics 2017;31:69-80.
- 643 <https://doi.org/10.1016/j.idc.2016.10.006>
- 644 Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma'ayan,
- 645 A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.
- 646 BMC Bioinformatics 2013;14:128. <https://doi.org/10.1186/1471-2105-14-128>
- 647 Chen, Y.; Ye, C.; Chen, J.; Lin, D.; Wang, H.; Wang, S. Association of the gene
- 648 polymorphisms of BMP2, ACVRL1, SMAD9 and their interactions with the risk of
- 649 essential hypertension in the Chinese Han population. Biosci Rep
- 650 2019;39:BSR20181217. <https://doi.org/10.1042/BSR20181217>
- 651 Chen, Z.; Lv, Y.; He, L.; Wu, S.; Wu, Z. Decreased CDKL2 Expression in Clear Cell Renal
- 652 Cell Carcinoma Predicts Worse Overall Survival. Front Mol Biosci 2021;8:657672.

653 <https://doi.org/10.3389/fmolb.2021.657672>

654 Fan, C.-T.; Lin, J.-C.; Lee, C.-H. Taiwan Biobank: a project aiming to aid Taiwan's

655 transition into a biomedical island. *Pharmacogenomics* 2008;9:235-246.

656 <https://doi.org/10.2217/14622416.9.2.235>

657 Filho, M.M.; Ribeiro, H.B.; Paula, L.J.C.; Nishioka, S.A.D.; Tamaki, W.T.; Costa, R.;

658 Siqueira, S.r.F.; Kawakami, J.T.; Higuchi, M.L. Endocarditis Secondary to Microsporidia:

659 Giant Vegetation in a Pacemaker User. *Circulation* 2009;119:e386–e388.

660 <https://doi.org/10.1161/CIRCULATIONAHA.108.817312>

661 Gorini, F.; Bustaffa, E.; Chatzianagnostou, K.; Bianchi, F.; Vassalle, C. Hydrogen sulfide

662 and cardiovascular disease: Doubts, clues, and interpretation difficulties from studies

663 in geothermal areas. *Sci Total Environ* 2020;743:140818.

664 <https://doi.org/10.1016/j.scitotenv.2020.140818>

665 Guauque-Olarte, S.; Messika-Zeitoun, D.; Droit, A.; Lamontagne, M.;

666 Tremblay-Marchand, J.; Lavoie-Charland, E.; Gaudreault, N.; Arsenault, B.J.; Dubé,

667 M.-P.; Tardif, J.-C.; Body, S.C.; Seidman, J.G.; Boileau, C.; Mathieu, P.; Pibarot, P.; Bossé,

668 Y. Calcium Signaling Pathway Genes RUNX2 and CACNA1C Are Associated With

669 Calcific Aortic Valve Disease. *Circulation: Cardiovascular Genetics* 2015;8:812–822.

670 <https://doi.org/10.1161/CIRCGENETICS.115.001145>

671 Hsu, B.M.; Chen, C.H.; Wan, M.T.; Cheng, H.W. Legionella prevalence in hot spring

672 recreation areas of Taiwan. Water Res 2006;40:3267-3273.
673 <https://doi.org/10.1016/j.watres.2006.07.007>

674 Hunter, D.J. Gene-environment interactions in human diseases. Nat Rev Genet
675 2005;6:287-298. <https://doi.org/10.1038/nrg1578>

676 Hwang, S.H.; Choi, Y.H.; Paik, H.J.; Wee, W.R.; Kim, M.K.; Kim, D.H. Potential
677 Importance of Ozone in the Association Between Outdoor Air Pollution and Dry Eye
678 Disease in South Korea. JAMA Ophthalmol 2016;134:503-510.
679 <https://doi.org/10.1001/jamaophthalmol.2016.0139>

680 ITO, I.; NAITO, J.; KADOWAKI, S.; MISHIMA, M.; ISHIDA, T.; HONGO, T.; MA, L.; ISHII, Y.;
681 MATSUMOTO, T.; YAMAGUCHI, K. Hot Spring Bath and Legionella Pneumonia: an
682 Association Confirmed by Genomic Identification. Internal Medicine
683 2002;41:859-863. <https://doi.org/10.2169/internalmedicine.41.859>

684 Kosrirukvongs, P.; Ngowyutagon, P.; Pusuwan, P.; Koolvisoot, A.; Nilganuwong, S.
685 Prevalence of dry eye syndrome and Sjogren's syndrome in patients with rheumatoid
686 arthritis. J Med Assoc Thai 2012;95 Suppl 4:S61-69.

687 Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.;
688 Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; McDermott, M.G.; Monteiro,
689 C.D.; Gundersen, G.W.; Ma'ayan, A. Enrichr: a comprehensive gene set enrichment
690 analysis web server 2016 update. Nucleic Acids Research 2016;44:W90-W97.

691 <https://doi.org/10.1093/nar/gkw377>

692 Li, N.; Zhu, L.; Zhu, C.; Zhou, H.; Zheng, D.; Xu, G.; Shi, H.; Gao, J.; Li, A.J.; Wang, Z.;

693 Sun, L.; Li, X.; Shao, G. BMPR2 promoter methylation and its expression in valvular

694 heart disease complicated with pulmonary artery hypertension. *Aging (Albany NY)*

695 2021;13:24580-24604. <https://doi.org/10.18632/aging.203690>

696 Magne, J.; Pibarot, P.; Sengupta, P.P.; Donal, E.; Rosenhek, R.; Lancellotti, P.

697 Pulmonary hypertension in valvular disease: a comprehensive review on

698 pathophysiology to therapy from the HAVEC Group. *JACC Cardiovasc Imaging*

699 2015;8:83-99. <https://doi.org/10.1016/j.jcmg.2014.12.003>

700 McAllister, K.; Mechanic, L.E.; Amos, C.; Aschard, H.; Blair, I.A.; Chatterjee, N.; Conti,

701 D.; Gauderman, W.J.; Hsu, L.; Hutter, C.M.; Jankowska, M.M.; Kerr, J.; Kraft, P.;

702 Montgomery, S.B.; Mukherjee, B.; Papanicolaou, G.J.; Patel, C.J.; Ritchie, M.D.; Ritz,

703 B.R.; Thomas, D.C.; Wei, P.; Witte, J.S. Current Challenges and New Opportunities for

704 Gene-Environment Interaction Studies of Complex Diseases. *Am J Epidemiol*

705 2017;186:753-761. <https://doi.org/10.1093/aje/kwx227>

706 Nordstrom, D.K.; Ball, J.W.; McCleskey, R.B. Ground water to surface water: Chemistry

707 of thermal outflows in Yellowstone National Park. *Geothermal biology and*

708 *geochemistry in Yellowstone National Park* 2005:73-94.

709 Nuvolone, D.; Petri, D.; Biggeri, A.; Barbone, F.; Voller, F.; Nuvolone, D.; Petri, D.;

710 Biggeri, A.; Barbone, F.; Voller, F. Health effects associated with short-term exposure
711 to hydrogen sulfide from geothermal power plants: a case-crossover study in the
712 geothermal areas in Tuscany. *International Archives of Occupational and*
713 *Environmental Health* 2020;93:669-682.
714 <https://doi.org/10.1007/s00420-020-01522-9>

715 Nuvolone, D.; Petri, D.; Pepe, P.; Voller, F. Health effects associated with chronic
716 exposure to low-level hydrogen sulfide from geothermoelectric power plants. A
717 residential cohort study in the geothermal area of Mt. Amiata in Tuscany. *Sci Total*
718 *Environ* 2019;659:973-982. <https://doi.org/10.1016/j.scitotenv.2018.12.363>

719 Pflugfelder, S.C.; de Paiva, C.S. The Pathophysiology of Dry Eye Disease: What We
720 Know and Future Directions for Research. *Ophthalmology* 2017;124:S4-S13.
721 <https://doi.org/10.1016/j.opthta.2017.07.010>

722 Rothman, K.J.; Greenland, S. *Modern epidemiology*. 2nd ed. Philadelphia, PA:
723 Lippincott-Raven; 1998

724 Serbulea, M.; Payyappallimana, U. Onsen (hot springs) in Japan—Transforming
725 terrain into healing landscapes. *Health & place*
726 2012;18 <https://doi.org/10.1016/j.healthplace.2012.06.020>

727 Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic,
728 M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; Jensen, L.J.; Mering, Christian v. STRING v11:

729 protein–protein association networks with increased coverage, supporting functional
730 discovery in genome-wide experimental datasets. *Nucleic Acids Research*
731 2018;47:D607-D613. <https://doi.org/10.1093/nar/gky1131>

732 Thompson, W.D. Effect Modification and the Limits of Biological Inference from
733 Epidemiologic Data. *J Clin Epidemiol* 1991;44:221-232. [https://doi.org/Doi](https://doi.org/10.1016/0895-4356(91)90033-6)
734 10.1016/0895-4356(91)90033-6

735 Tracey, L.; Villuendas, R.; Dotor, A.M.; Spiteri, I.; Ortiz, P.; Garcia, J.F.; Peralto, J.L.;
736 Lawler, M.; Piris, M.A. Mycosis fungoides shows concurrent deregulation of multiple
737 genes involved in the TNF signaling pathway: an expression profile study. *Blood*
738 2003;102:1042-1050. <https://doi.org/10.1182/blood-2002-11-3574>

739 Tsubota, K.; Yokoi, N.; Shimazaki, J.; Watanabe, H.; Dogru, M.; Yamada, M.; Kinoshita,
740 S.; Kim, H.M.; Tchah, H.W.; Hyon, J.Y.; Yoon, K.C.; Seo, K.Y.; Sun, X.; Chen, W.; Liang, L.;
741 Li, M.; Liu, Z.; Asia Dry Eye, S. New Perspectives on Dry Eye Definition and Diagnosis:
742 A Consensus Report by the Asia Dry Eye Society. *Ocul Surf* 2017;15:65-76.
743 <https://doi.org/10.1016/j.jtos.2016.09.003>

744 Uchino, Y. The Ocular Surface Glycocalyx and its Alteration in Dry Eye Disease: A
745 Review. *Investigative Ophthalmology & Visual Science* 2018;59:DES158-DES162.
746 <https://doi.org/10.1167/iovs.17-23756>

747 Verhagen, A.P.; Bierma-Zeinstra, S.M.; Boers, M.; Cardoso, J.R.; Lambeck, J.; De Bie, R.;

748 De Vet, H.C. Balneotherapy (or spa therapy) for rheumatoid arthritis. An abridged
749 version of Cochrane Systematic Review. *Eur J Phys Rehabil Med* 2015;51:833-847.

750 Vickers, A.J.; van Calster, B.; Steyerberg, E.W. A simple, step-by-step guide to
751 interpreting decision curve analysis. *Diagnostic and Prognostic Research* 2019;3:18.
752 <https://doi.org/10.1186/s41512-019-0064-7>

753 Virolainen, S.J.; VonHandorf, A.; Viel, K.C.M.F.; Weirauch, M.T.; Kottyan, L.C.;
754 Virolainen, S.J.; VonHandorf, A.; Viel, K.C.M.F.; Weirauch, M.T.; Kottyan, L.C.
755 Gene-environment interactions and their impact on human health. *Genes &*
756 *Immunity* 2022;24:1-11. <https://doi.org/10.1038/s41435-022-00192-6>

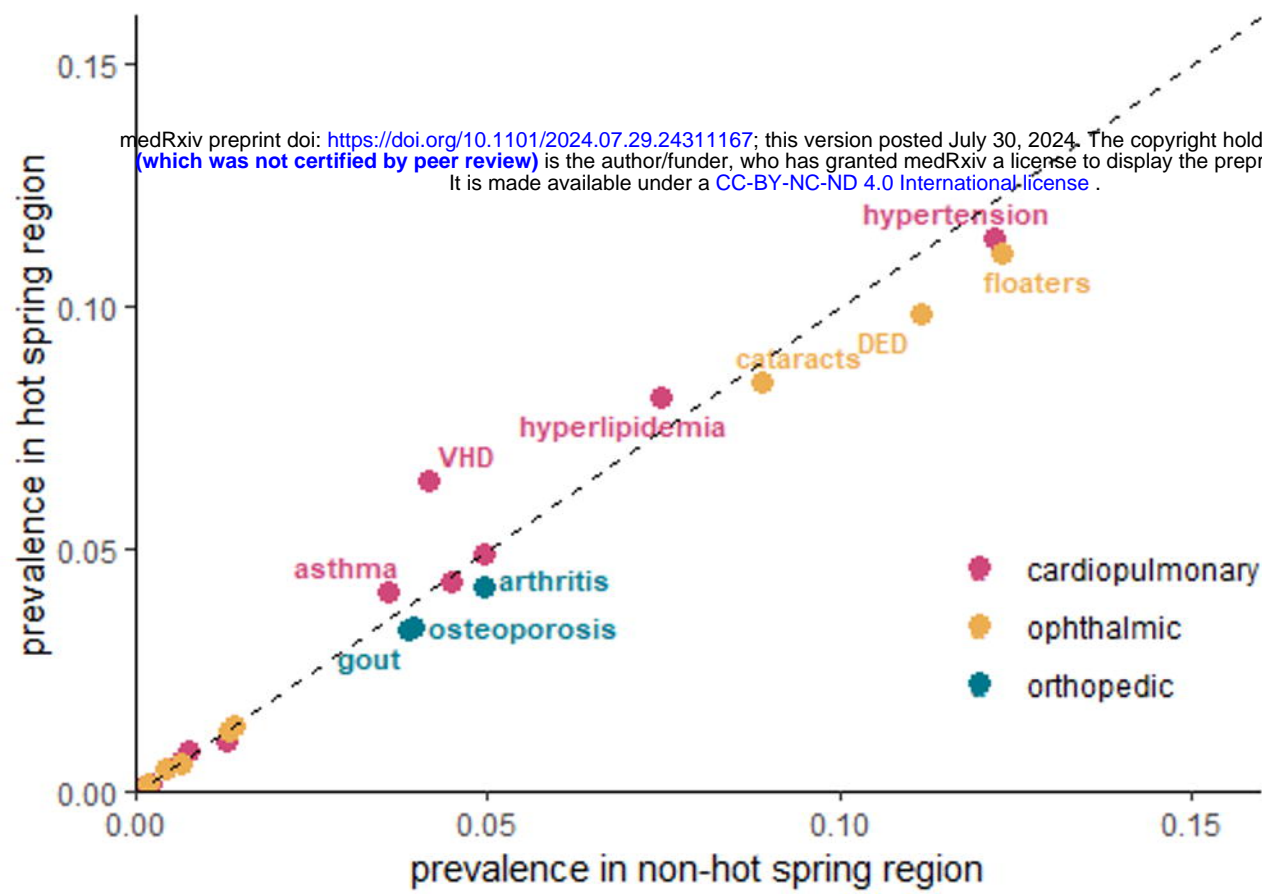
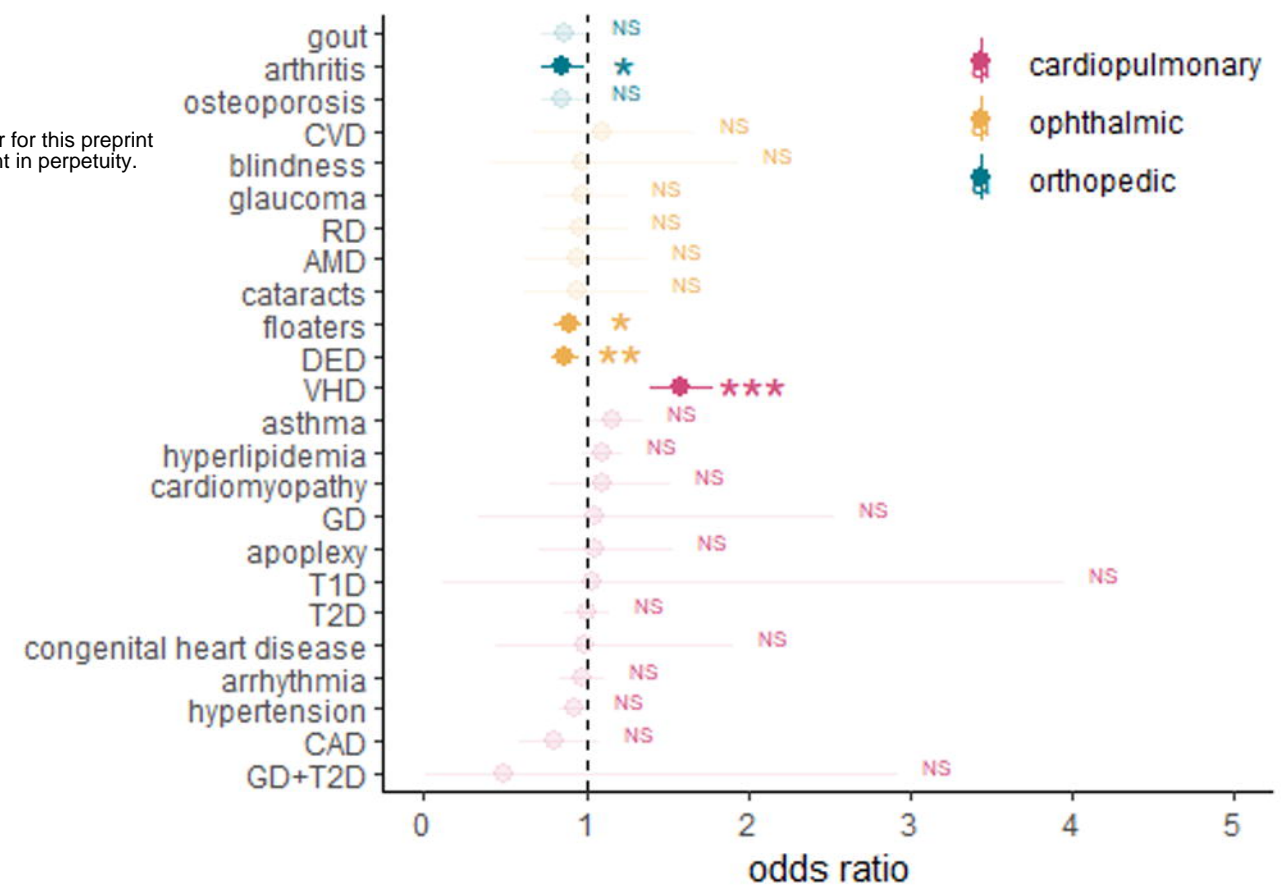
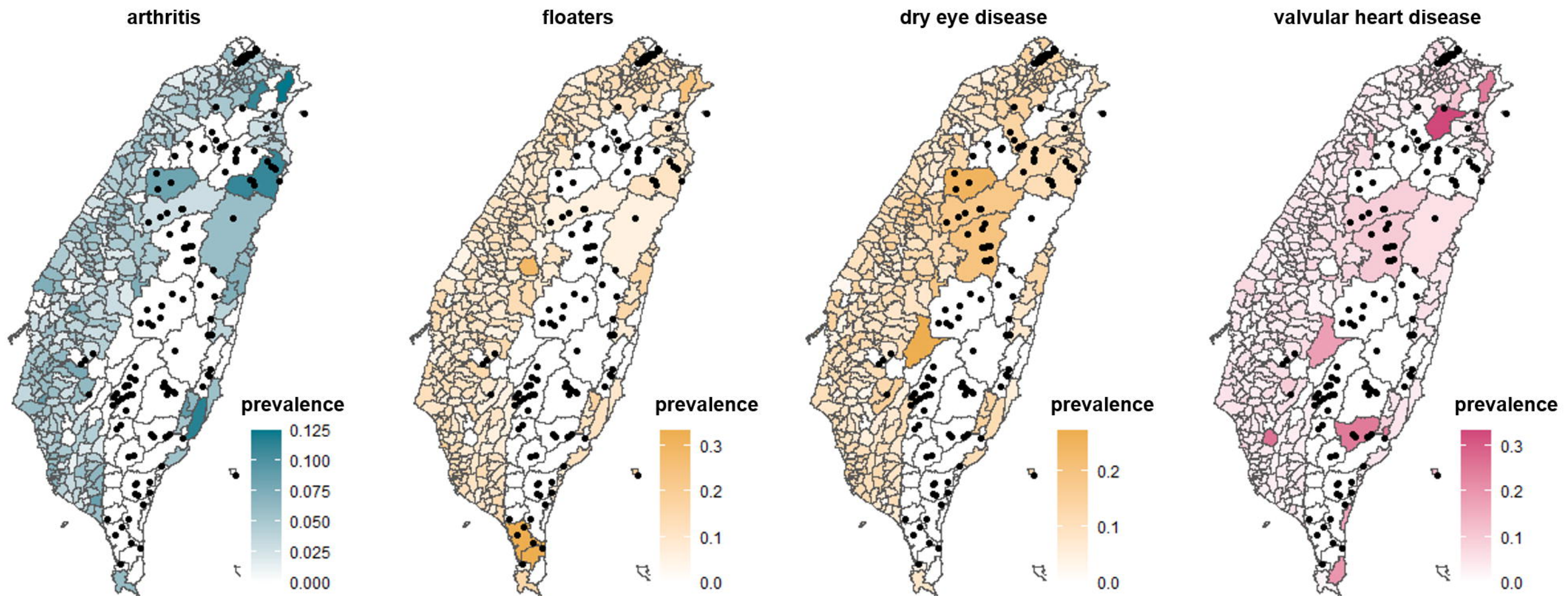
757 Xia, L.-y.; Luo, Z.; Xu, H.-p.; Liu, L.; Huai, W.-y.; Xia, J.; Yin, Q.-z.; Zhang, T.-e.; Chen, Y.-h.
758 Decoding Hippophae rhamnoides' action of mechanism in Sjögren's syndrome: A
759 network pharmacology and molecular docking study. *Quality Assurance and Safety of*
760 *Crops & Foods* 2023;15:207-250. <https://doi.org/10.15586/gas.v15i1.1210>

761 Xie, Z.; Bailey, A.; Kuleshov, M.V.; Clarke, D.J.B.; Evangelista, J.E.; Jenkins, S.L.;
762 Lachmann, A.; Wojciechowicz, M.L.; Kropiwnicki, E.; Jagodnik, K.M.; Jeon, M.;
763 Ma'ayan, A. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* 2021;1:e90.
764 <https://doi.org/10.1002/cpz1.90>

765 Yang, Q.; Khoury, M.J. Evolving methods in genetic epidemiology. III.
766 Gene-environment interaction in epidemiologic research. *Epidemiol Rev*

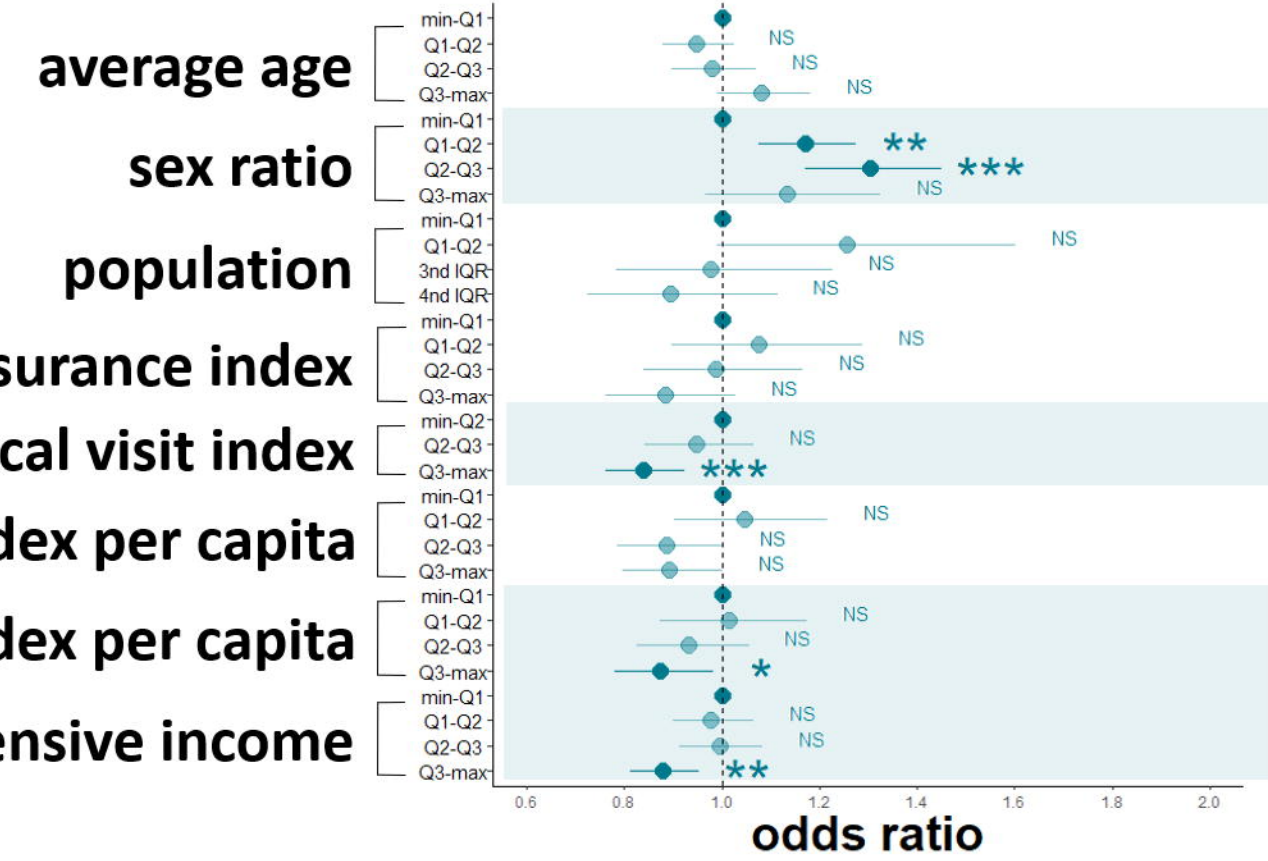
767 1997;19:33-43. <https://doi.org/10.1093/oxfordjournals.epirev.a017944>

768

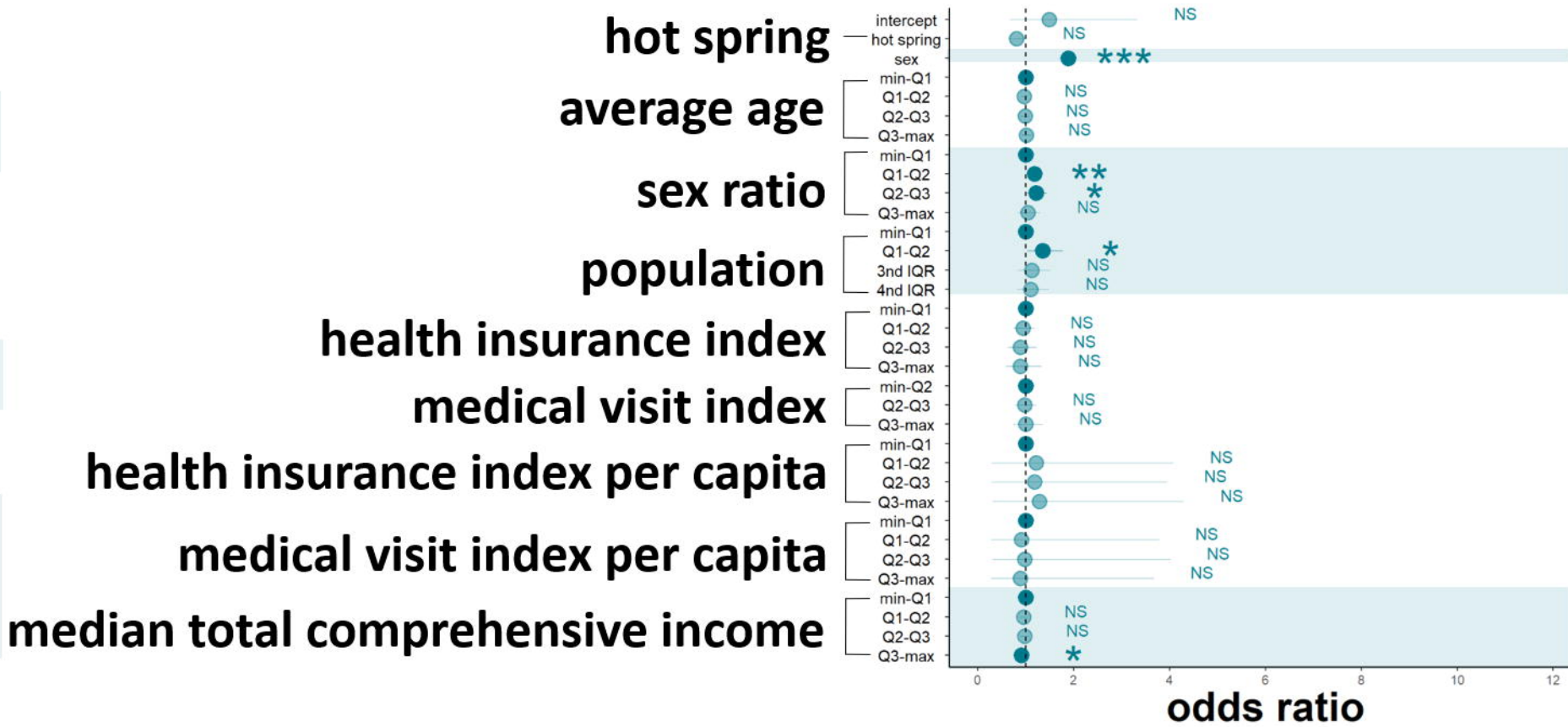
a**b****c**

arthritis

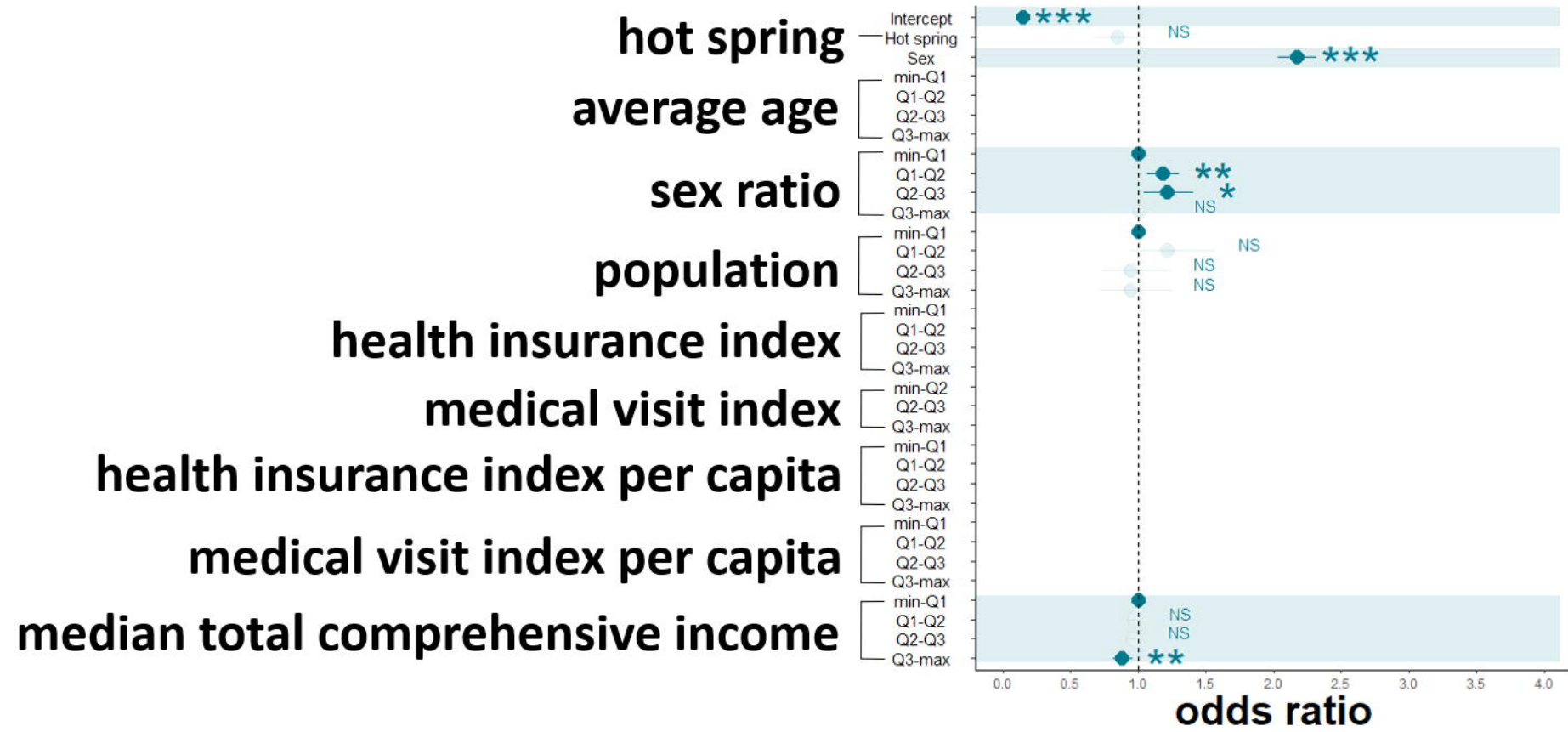
univariable



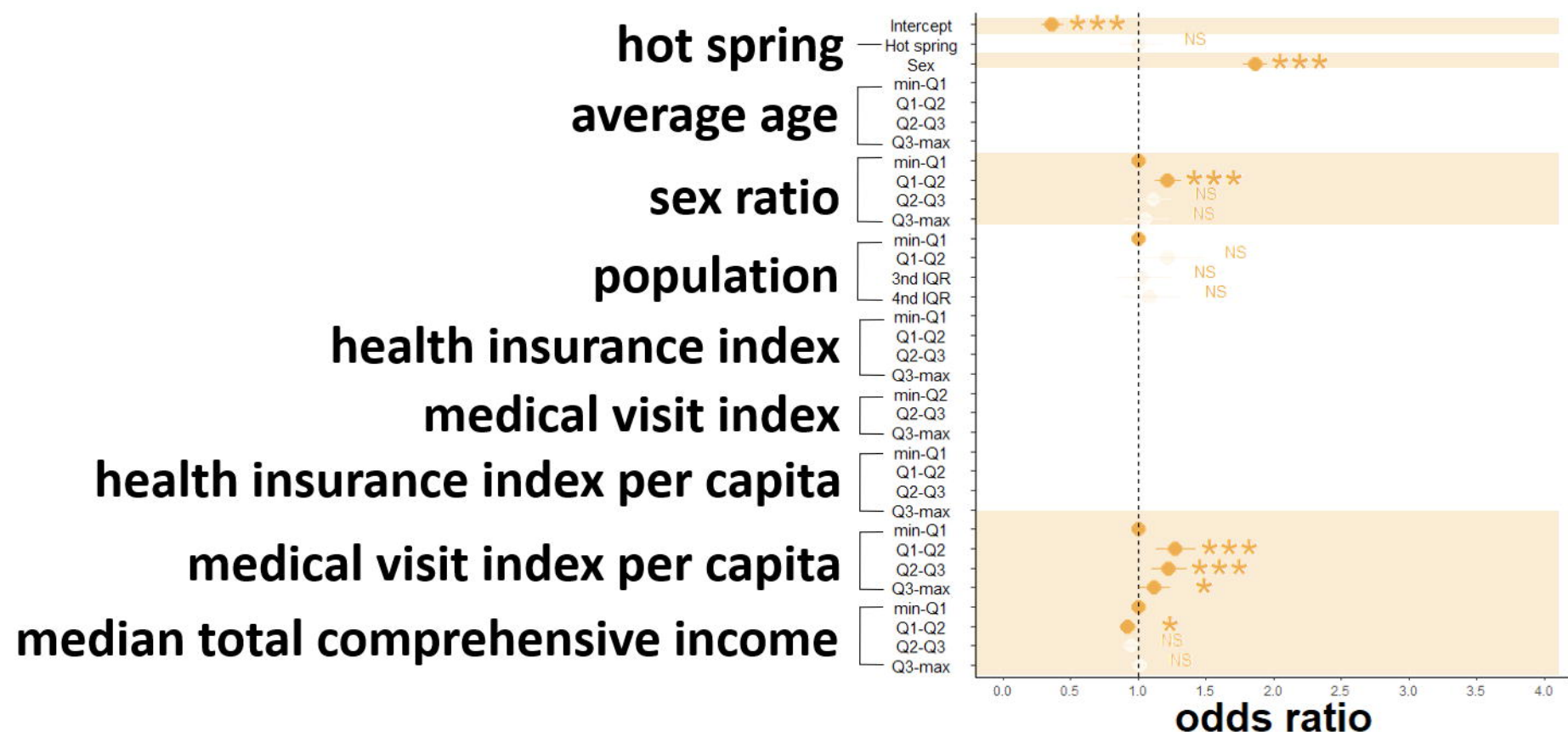
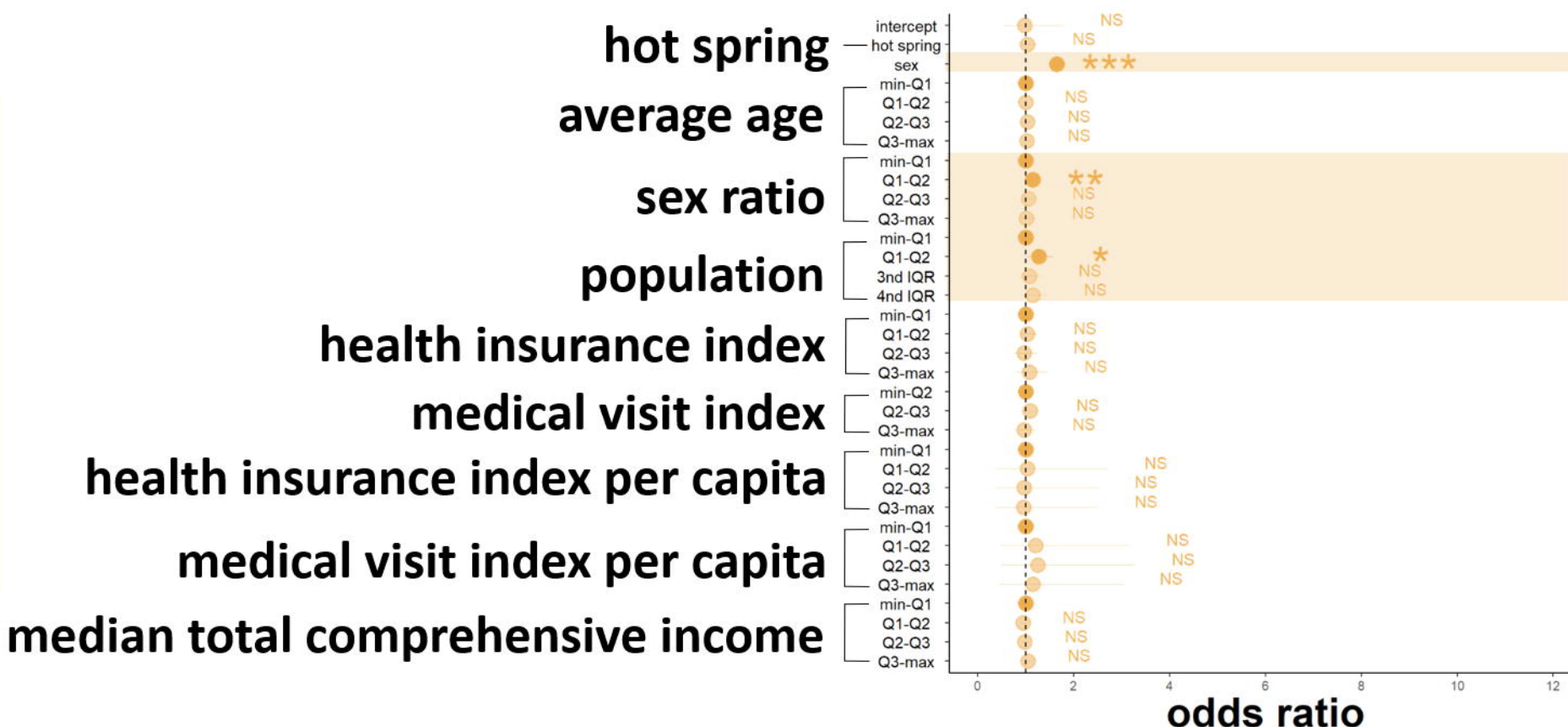
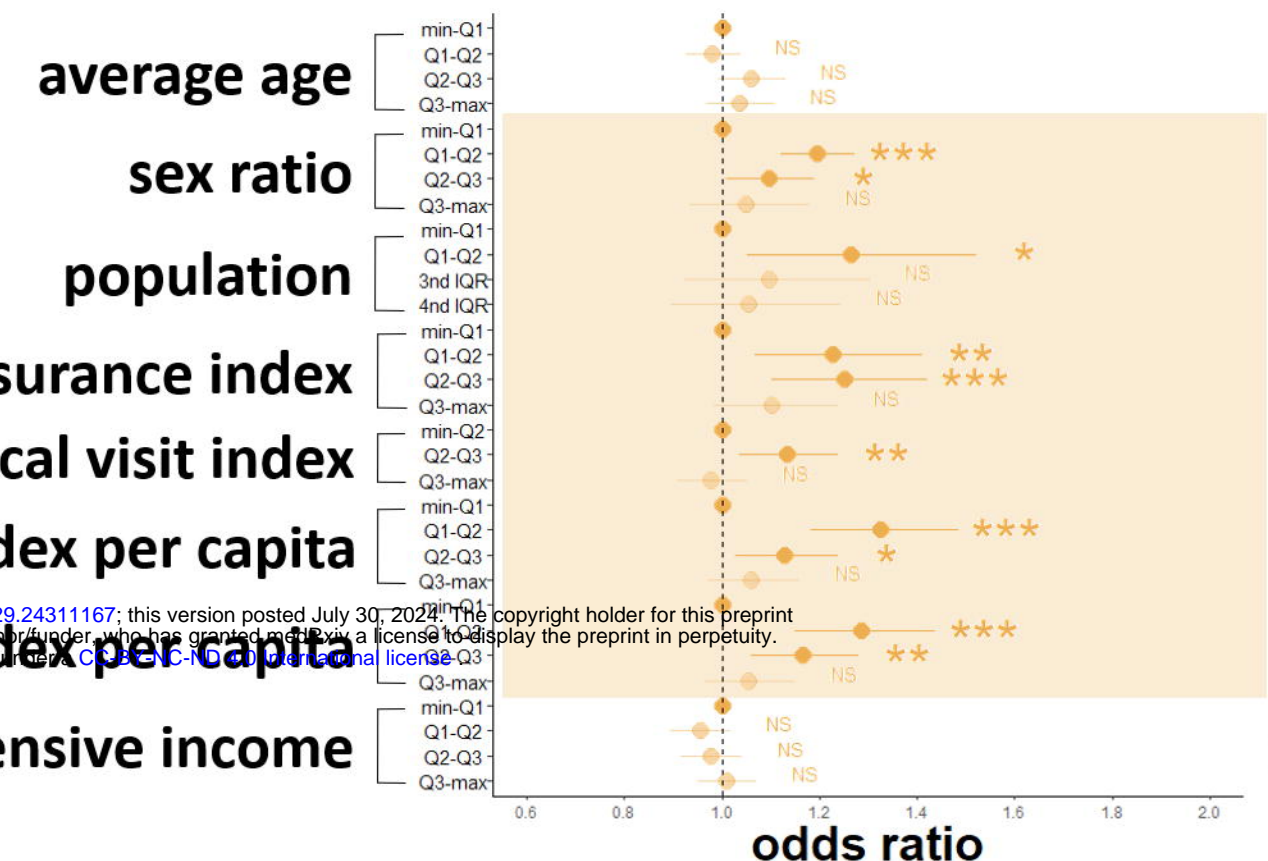
multivariable



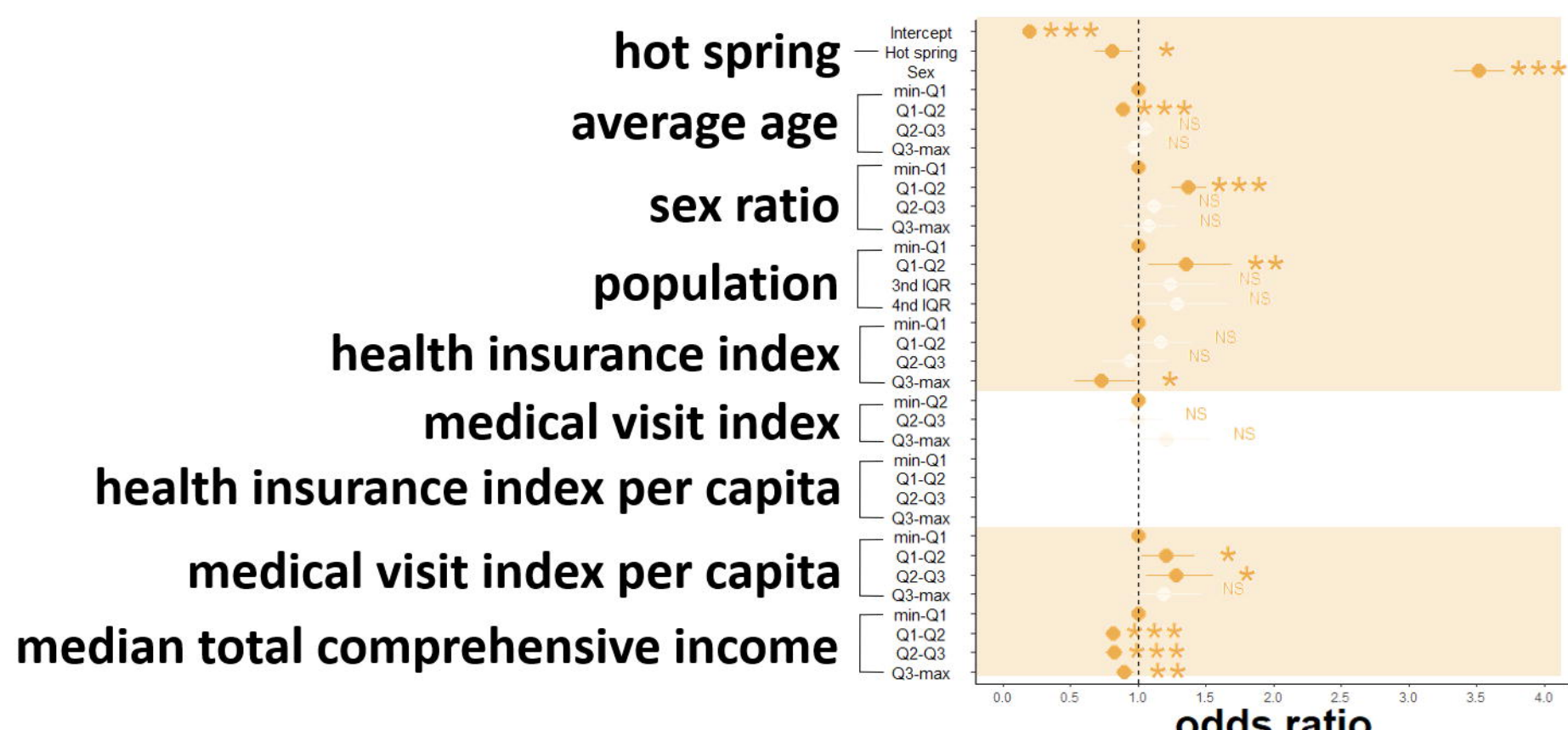
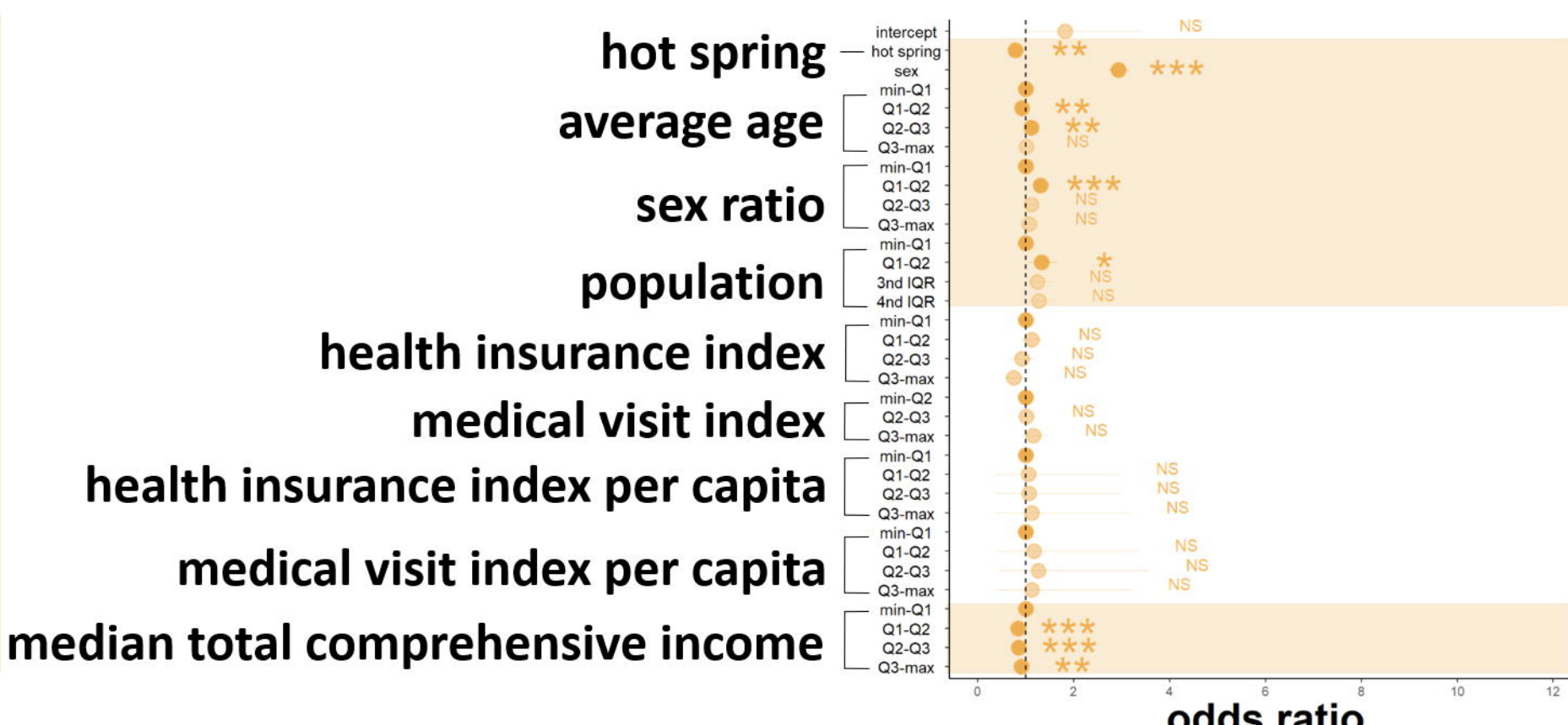
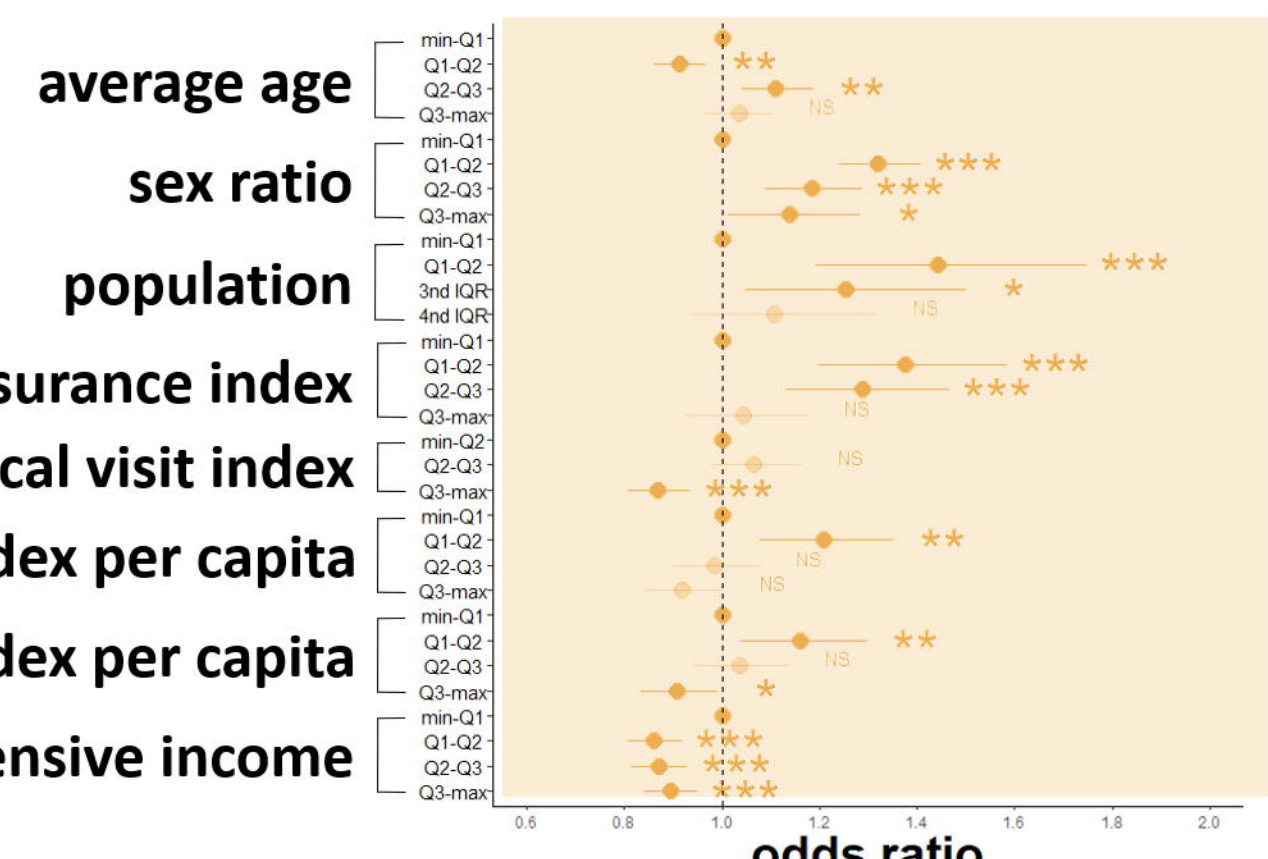
stepwise



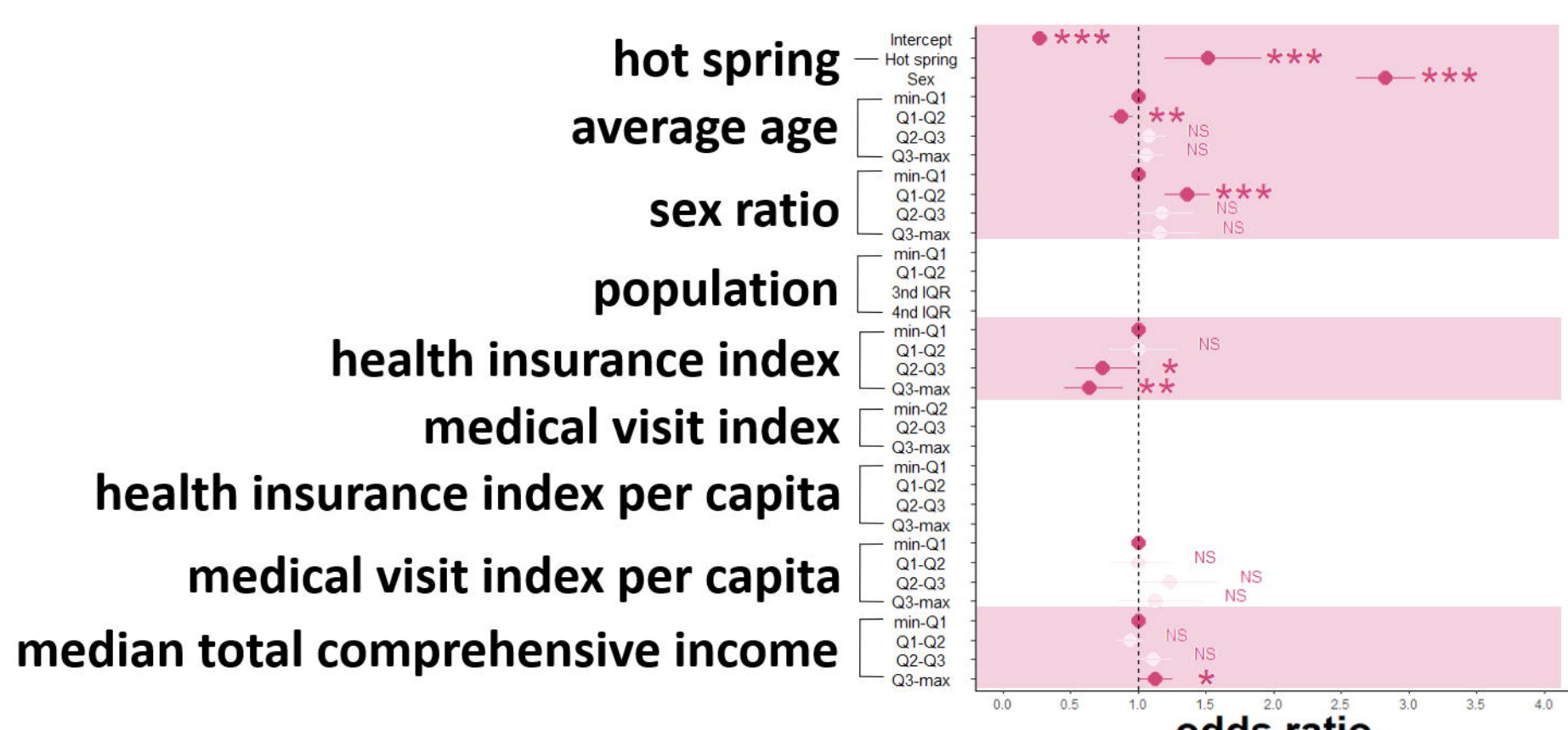
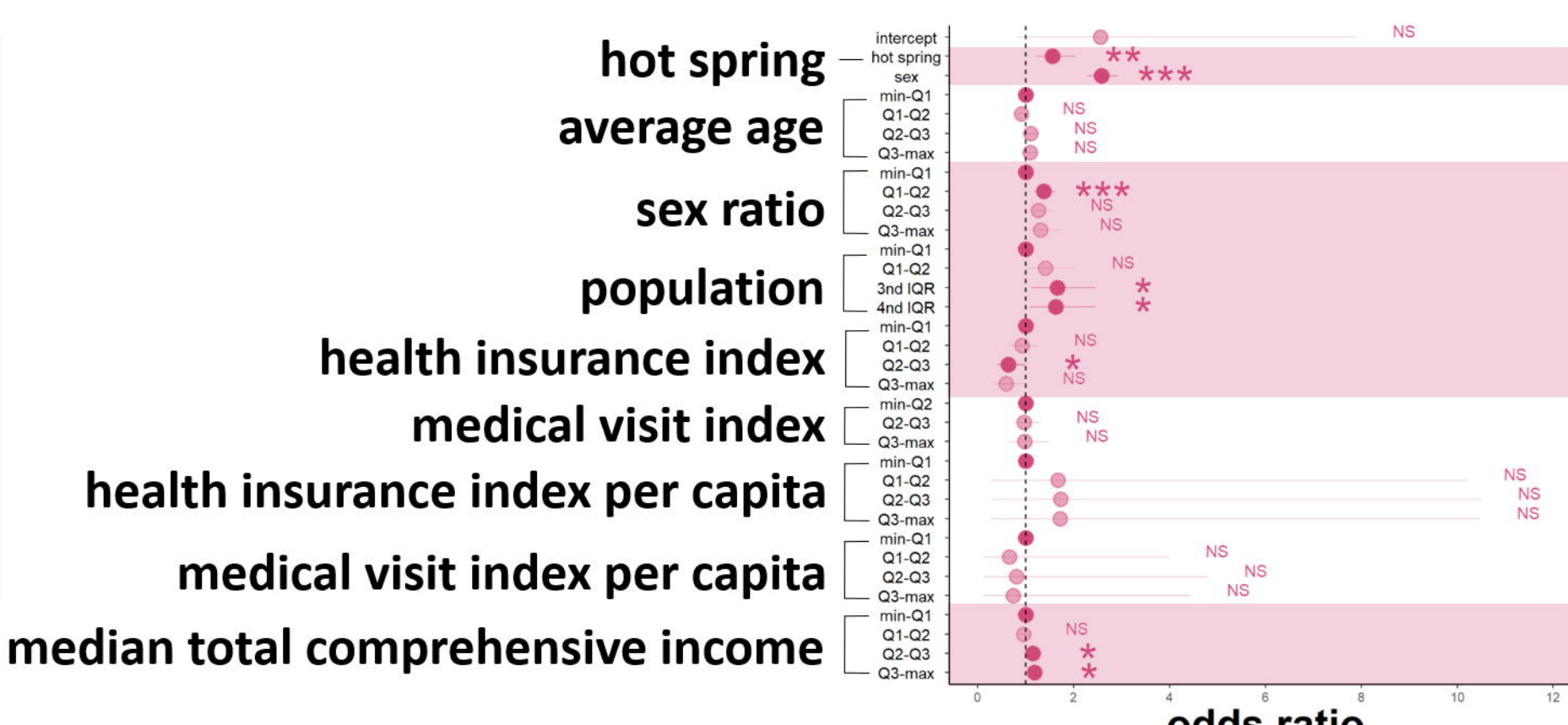
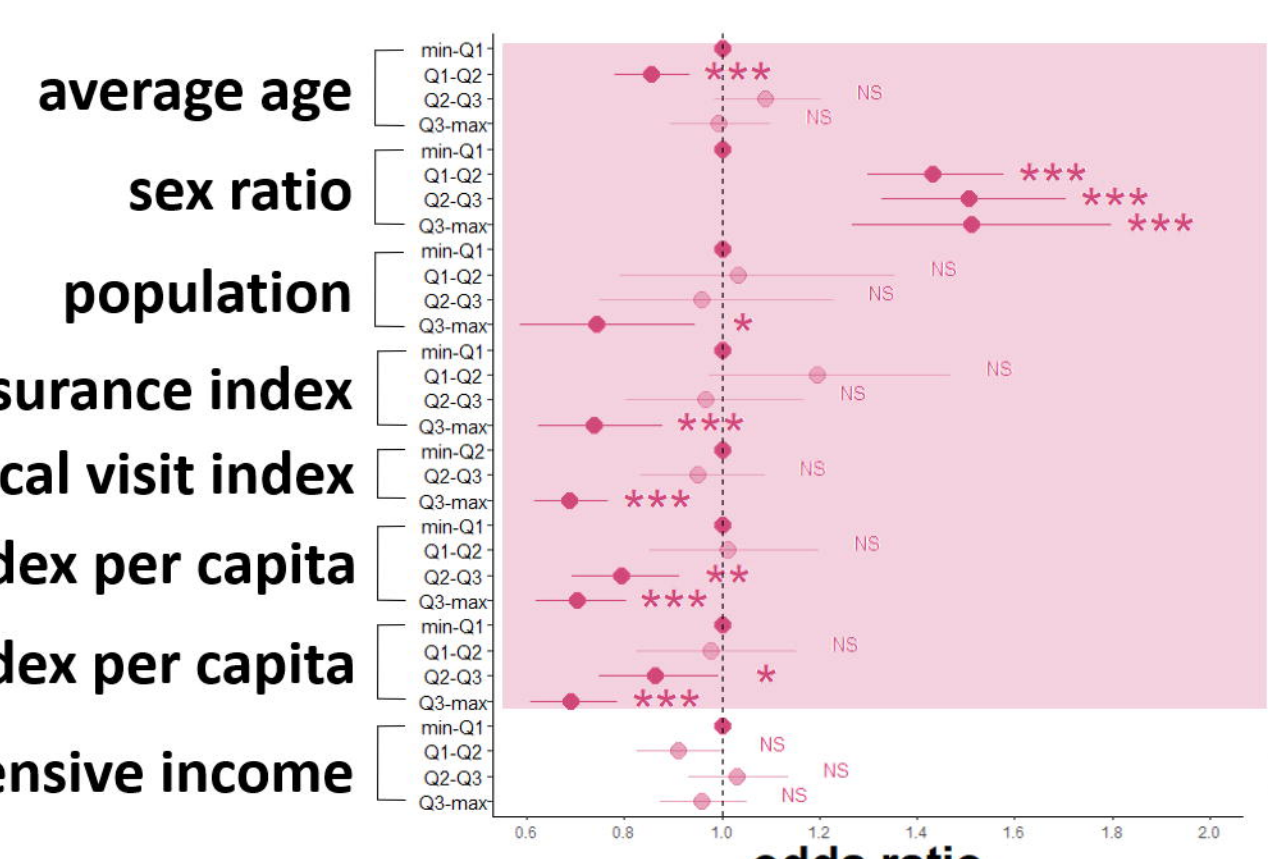
floaters



DED



VHD



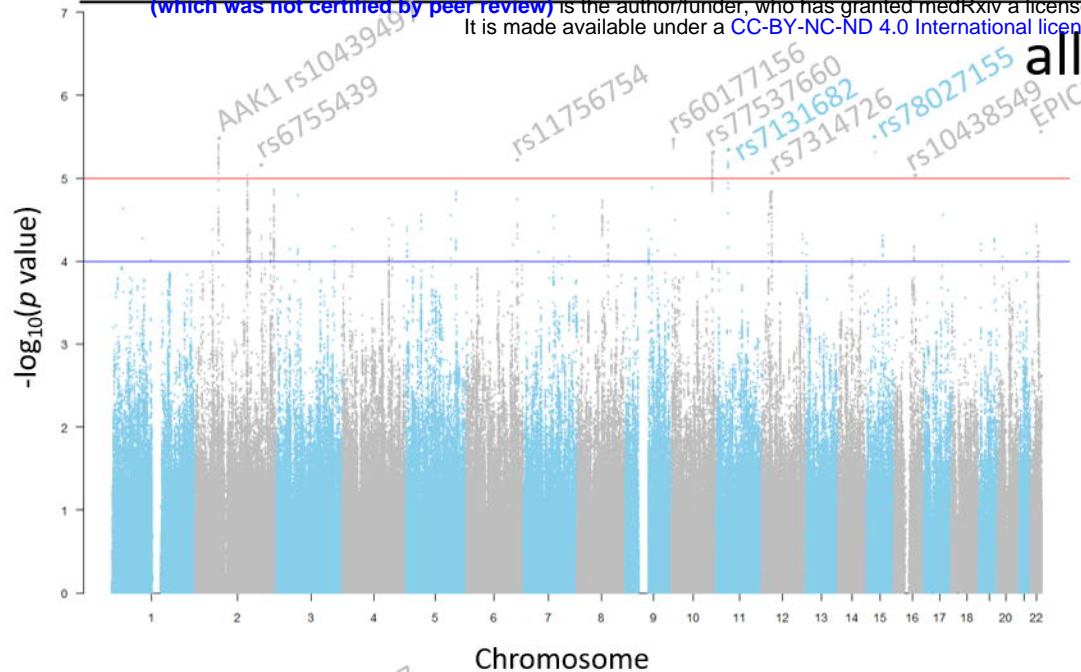
medRxiv preprint doi: <https://doi.org/10.1101/2024.07.29.24311167>; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dry eye disease

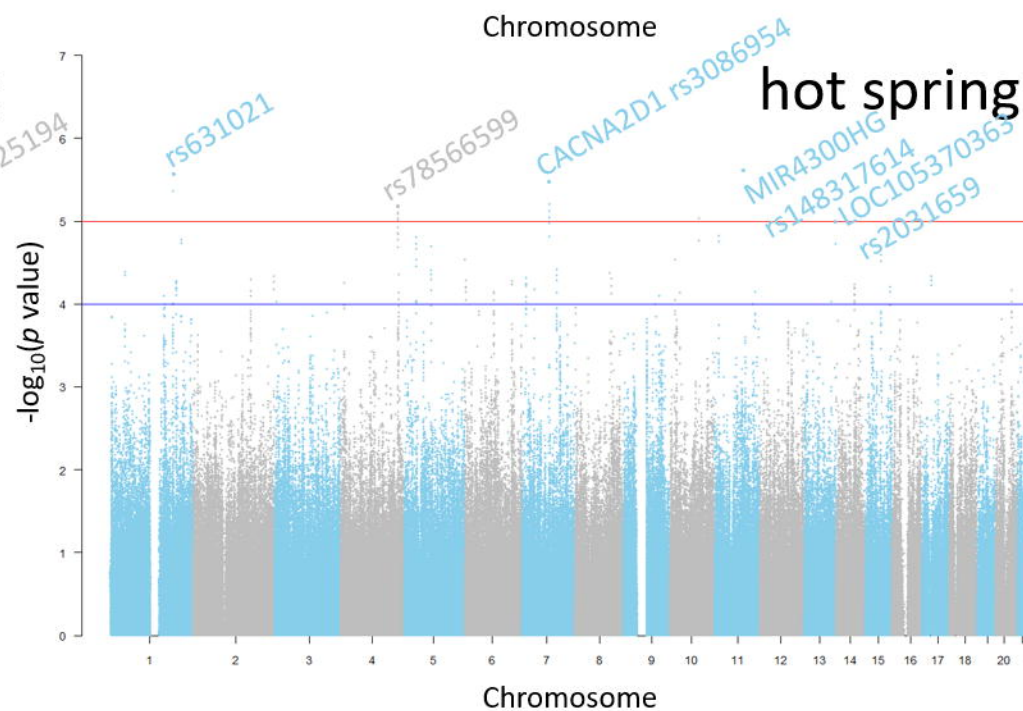
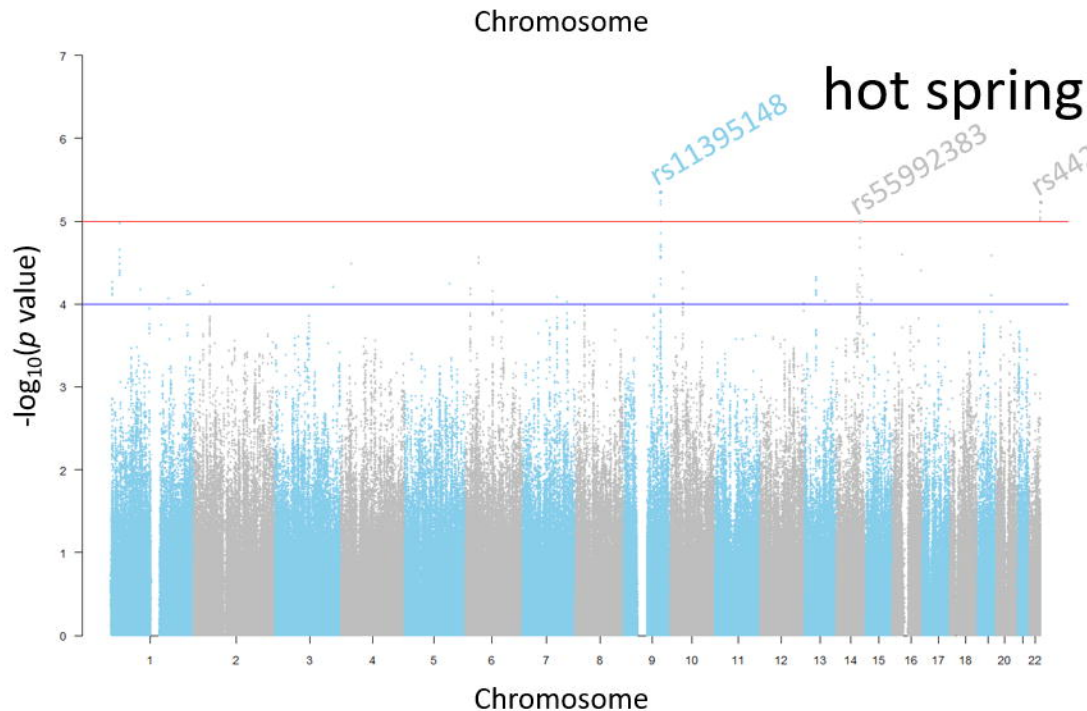
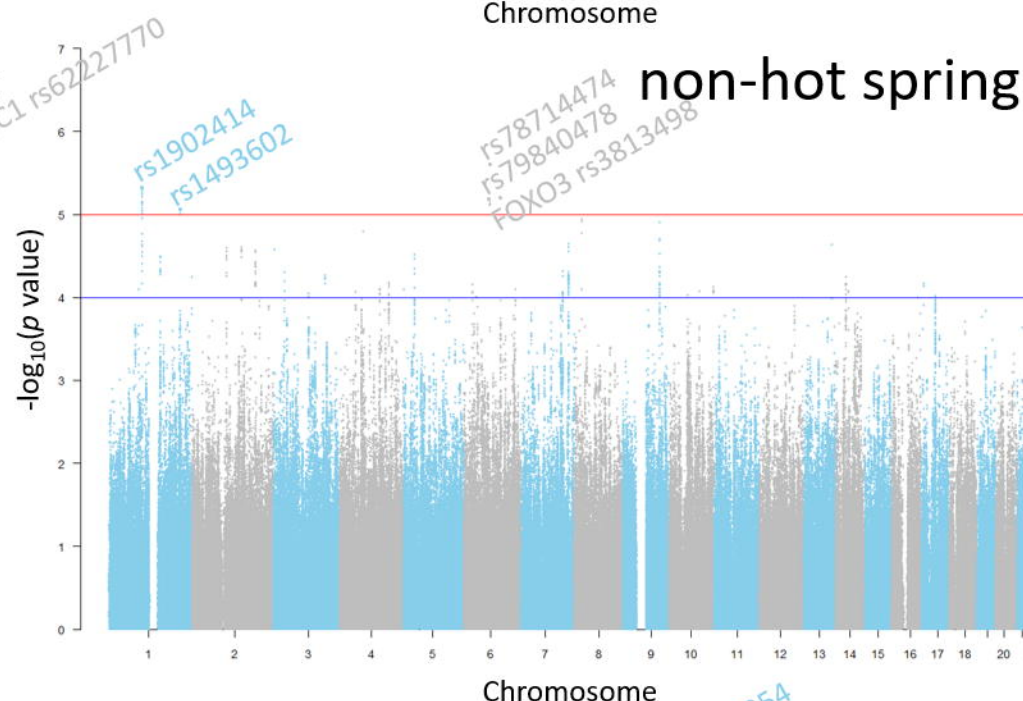
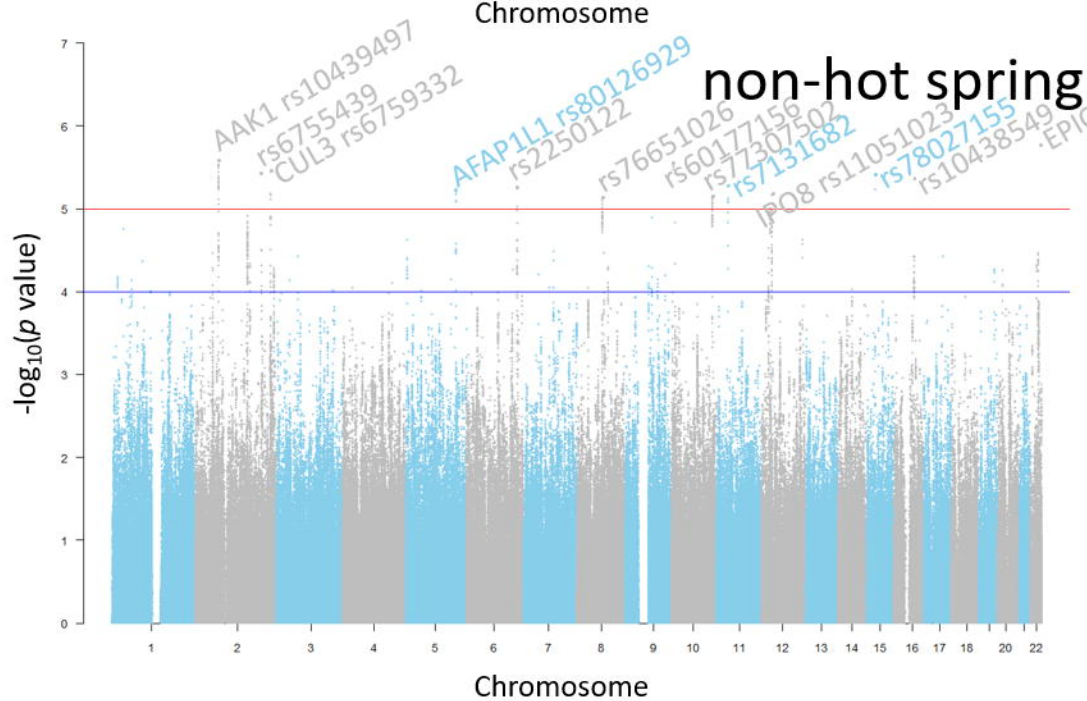
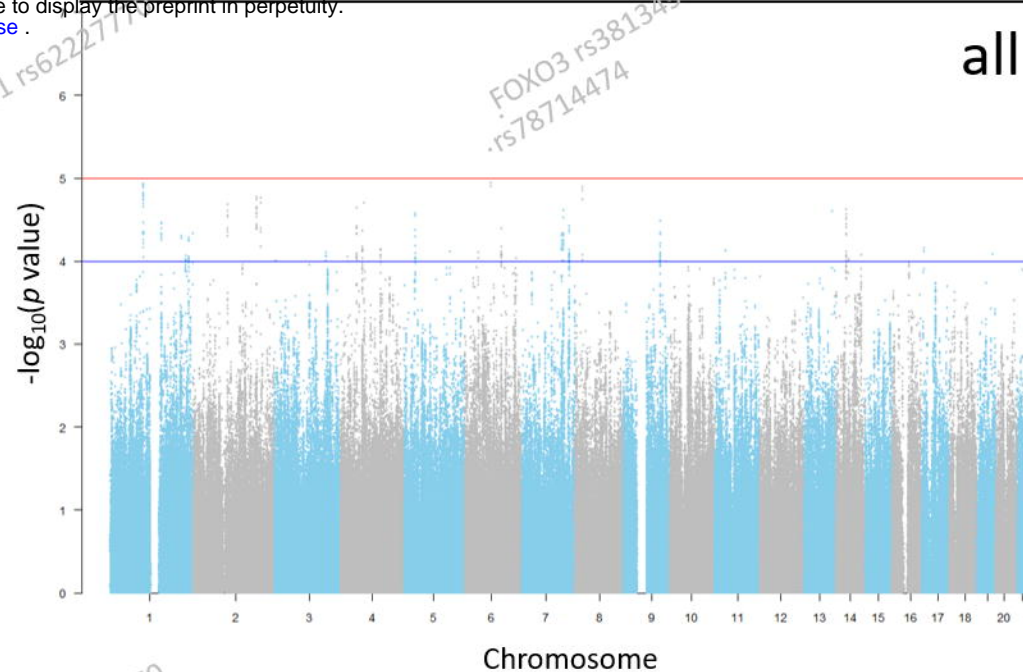
medRxiv preprint doi: <https://doi.org/10.1101/2024.07.29.24311167>; this version posted July 30, 2024. The copyright holder for this preprint

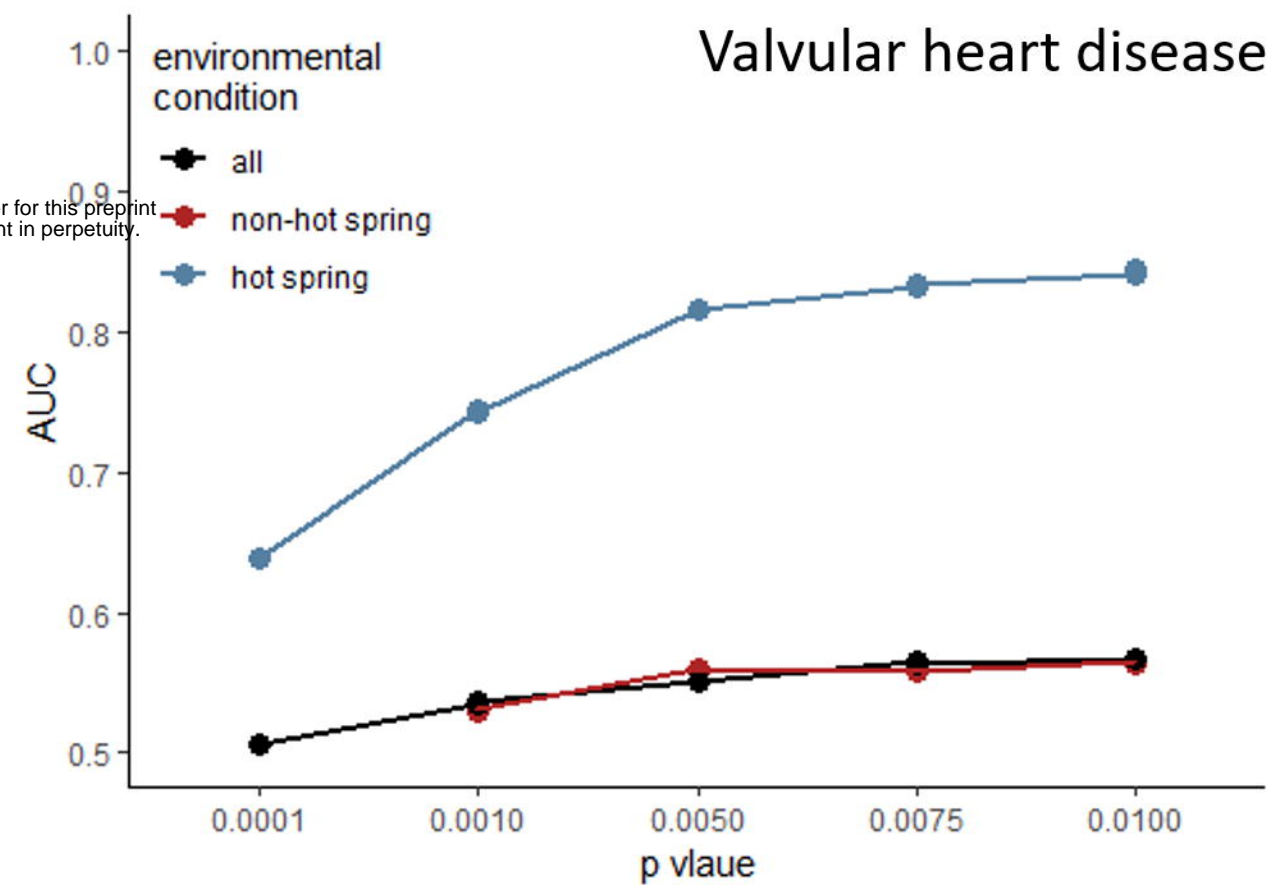
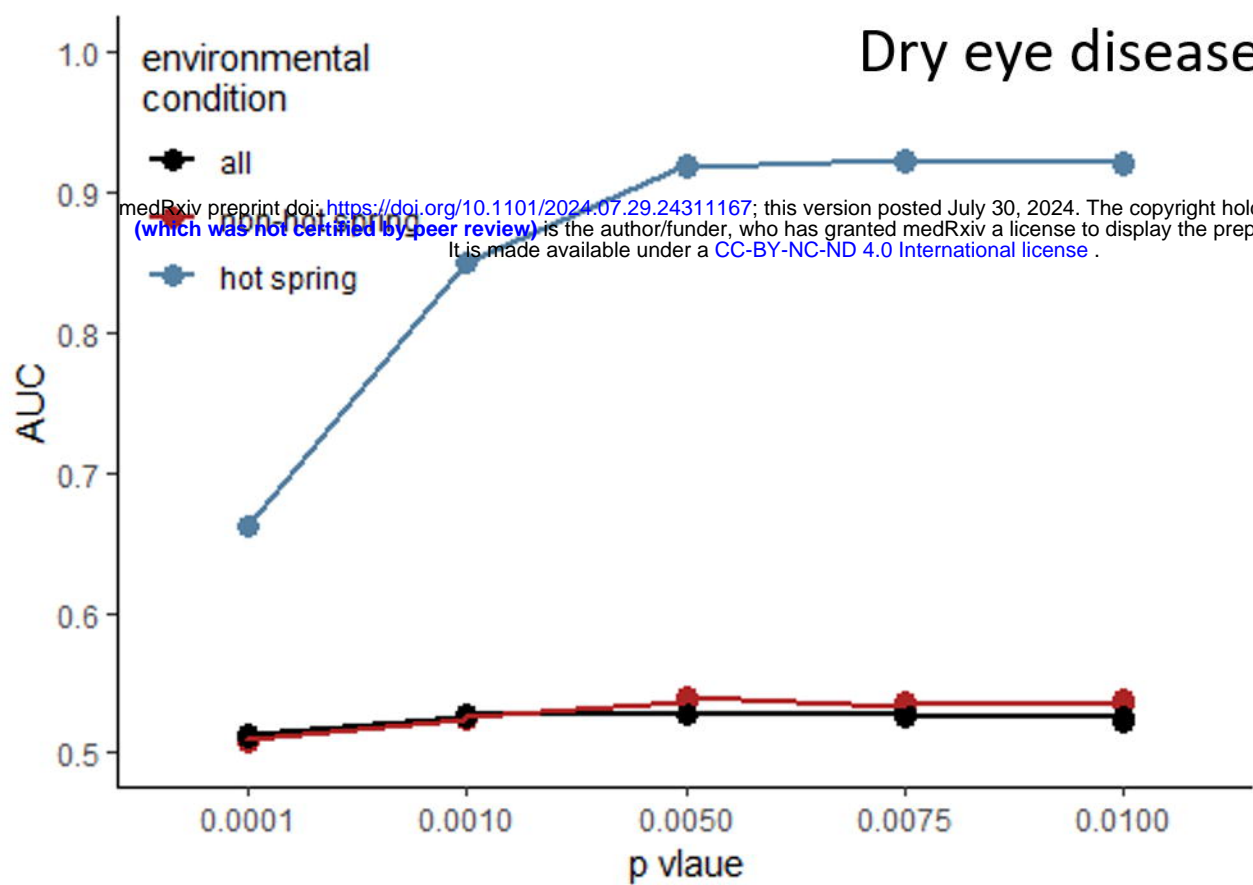
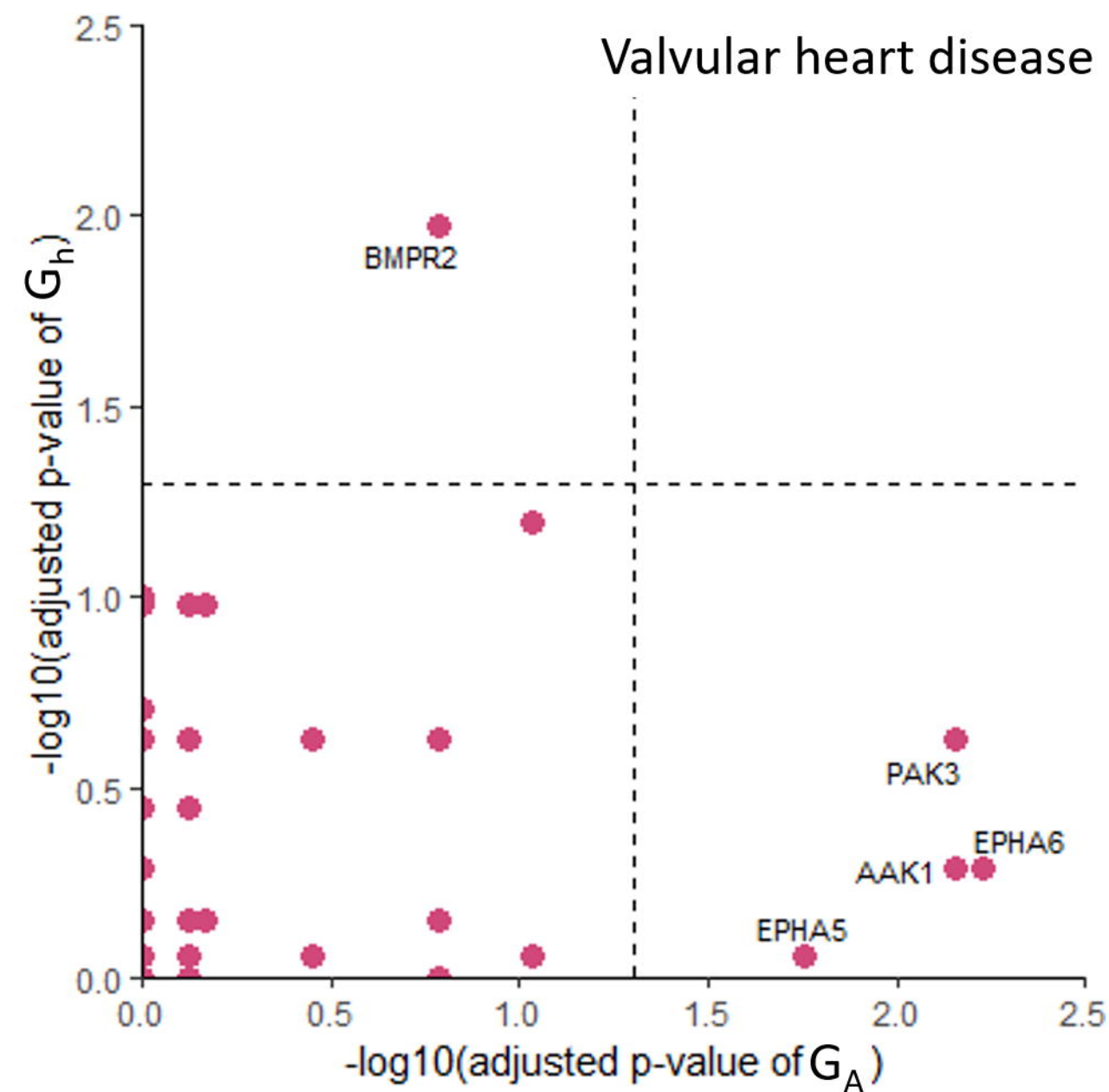
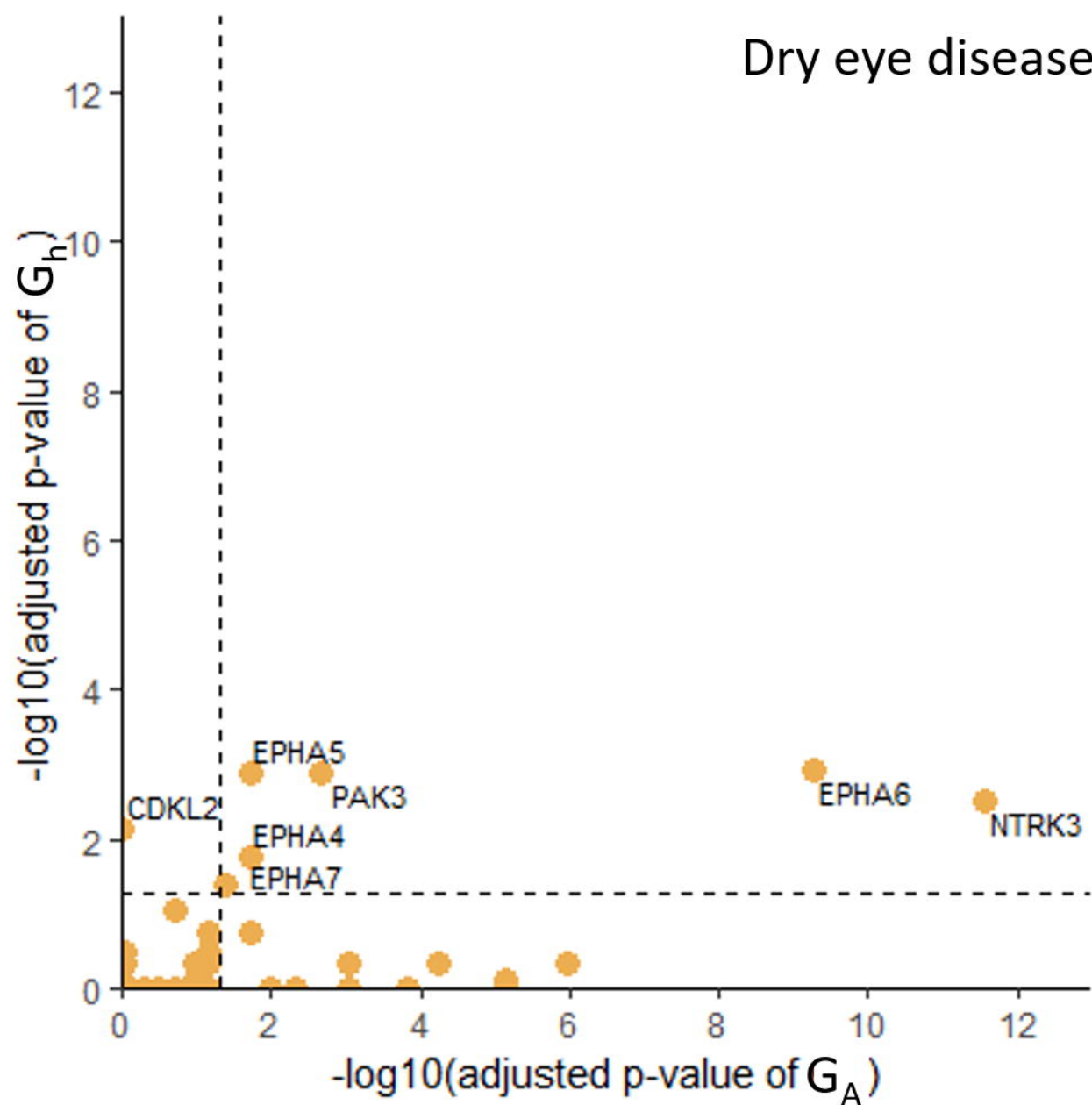
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

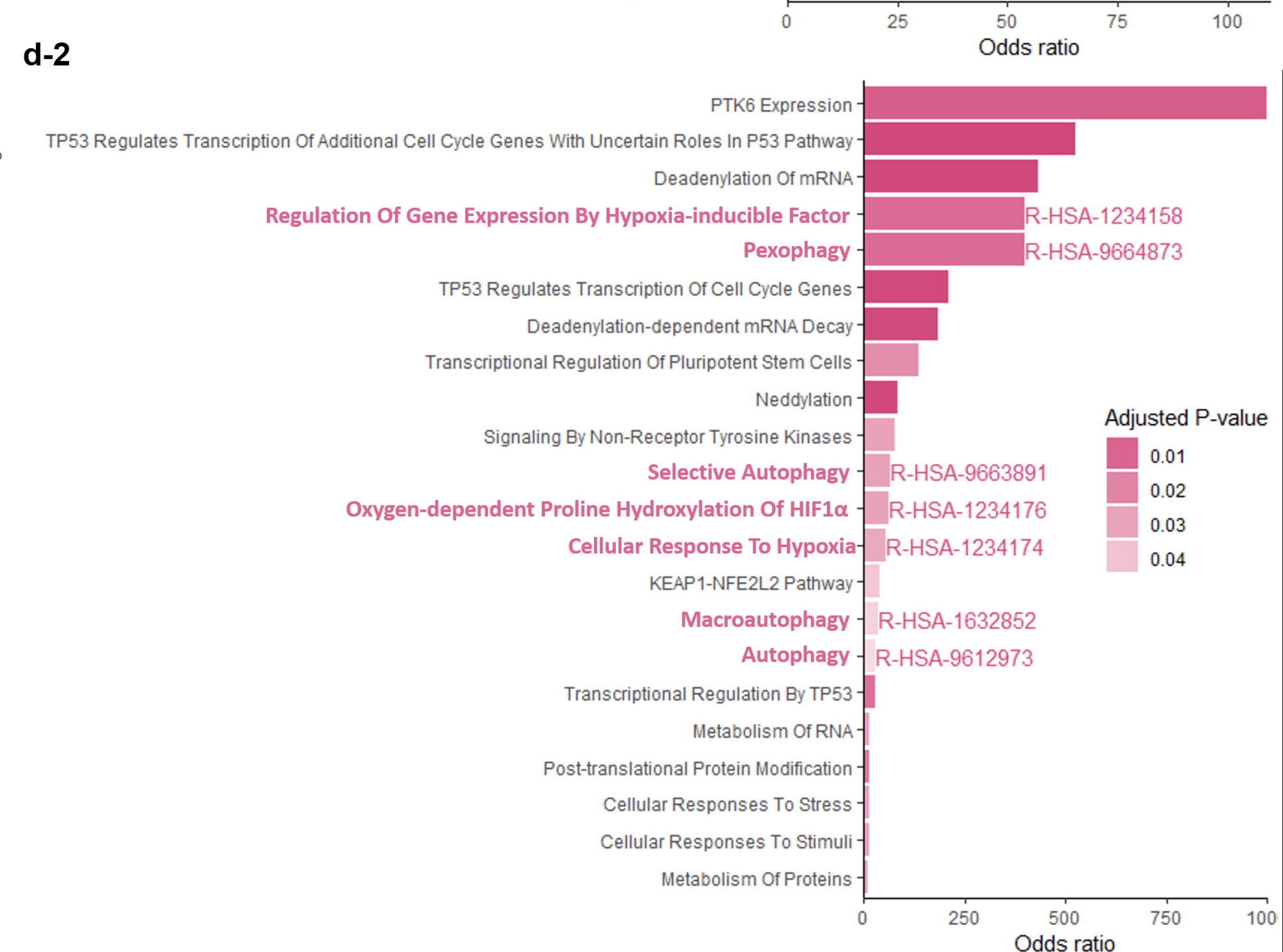
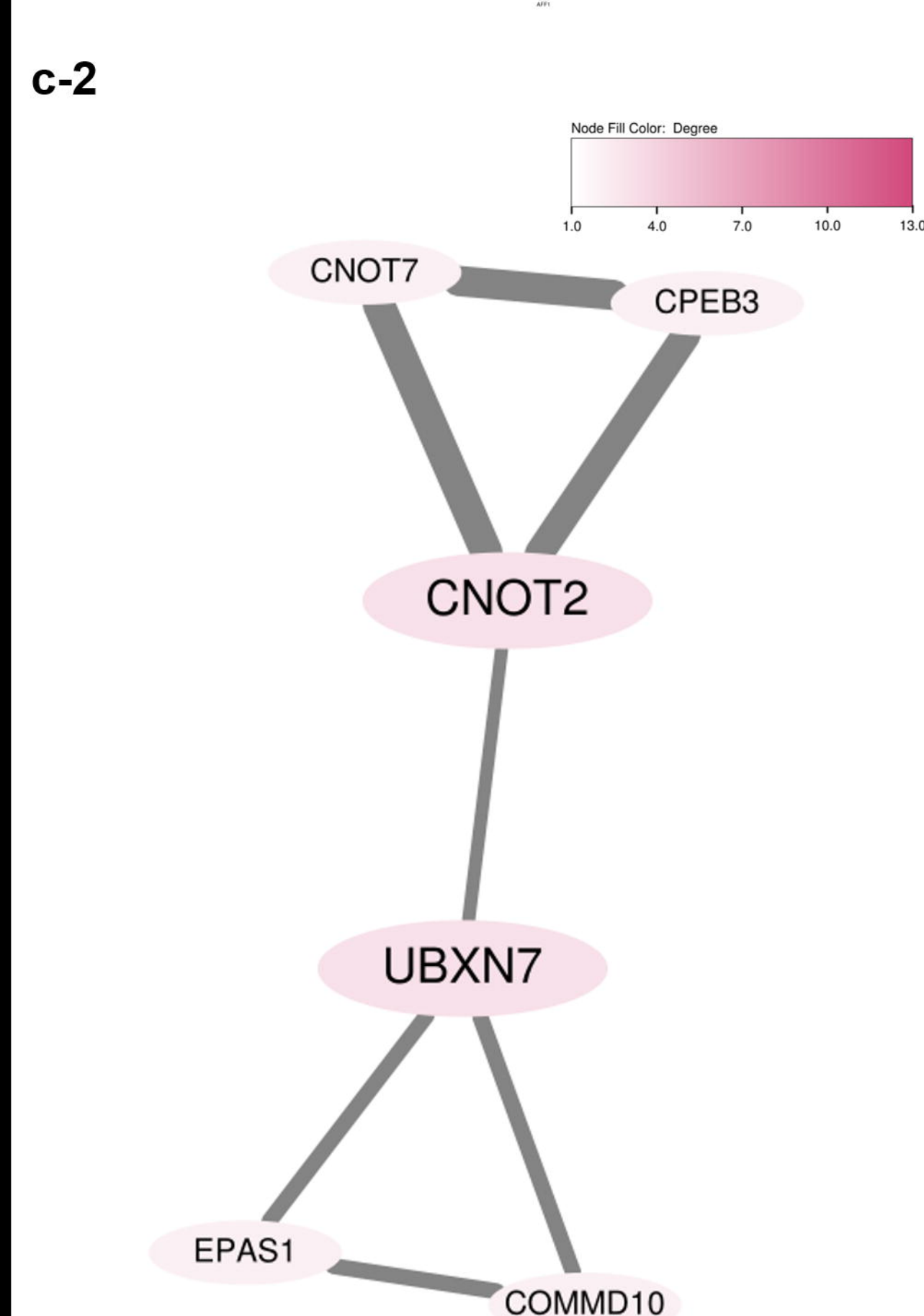
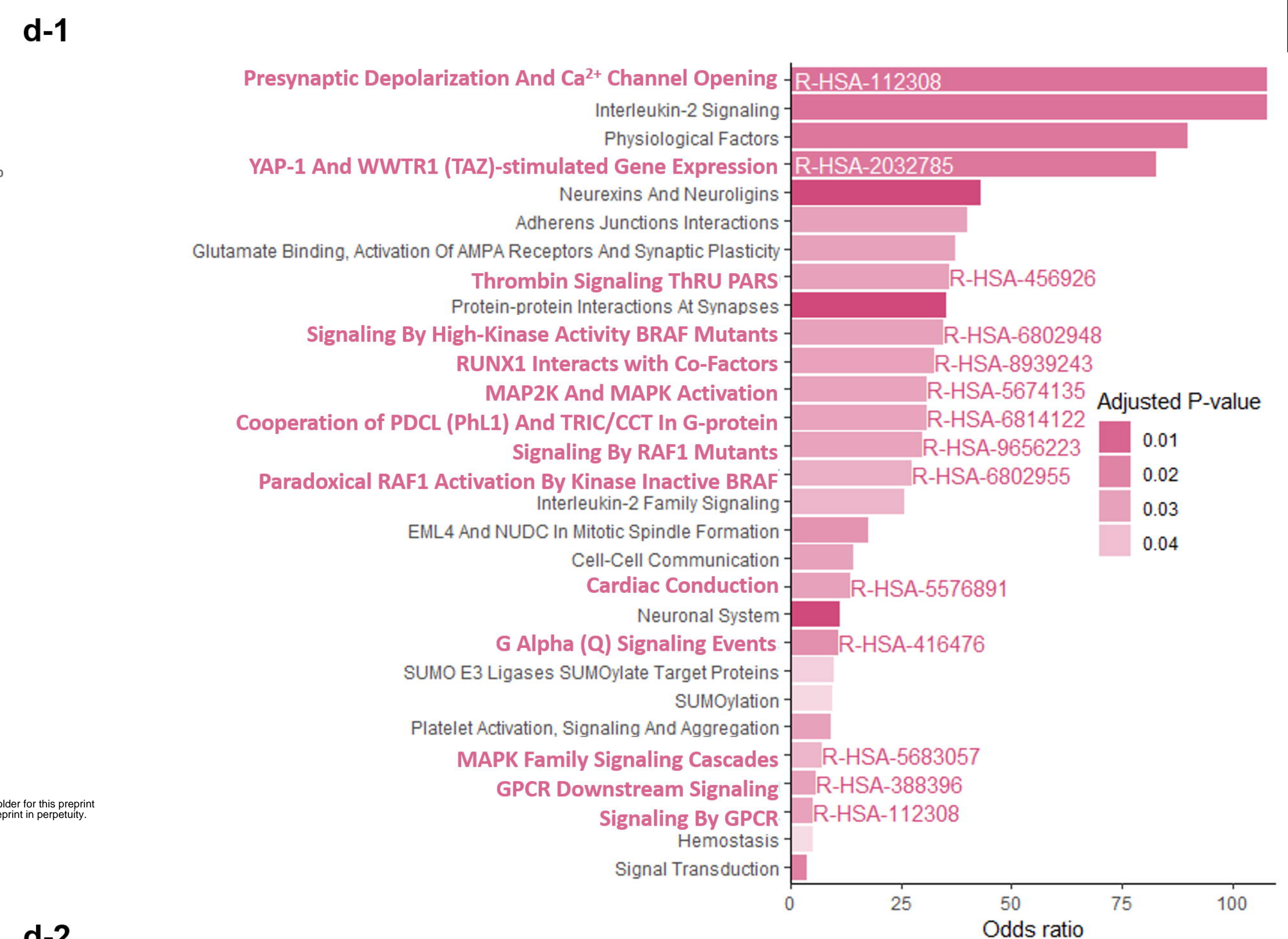
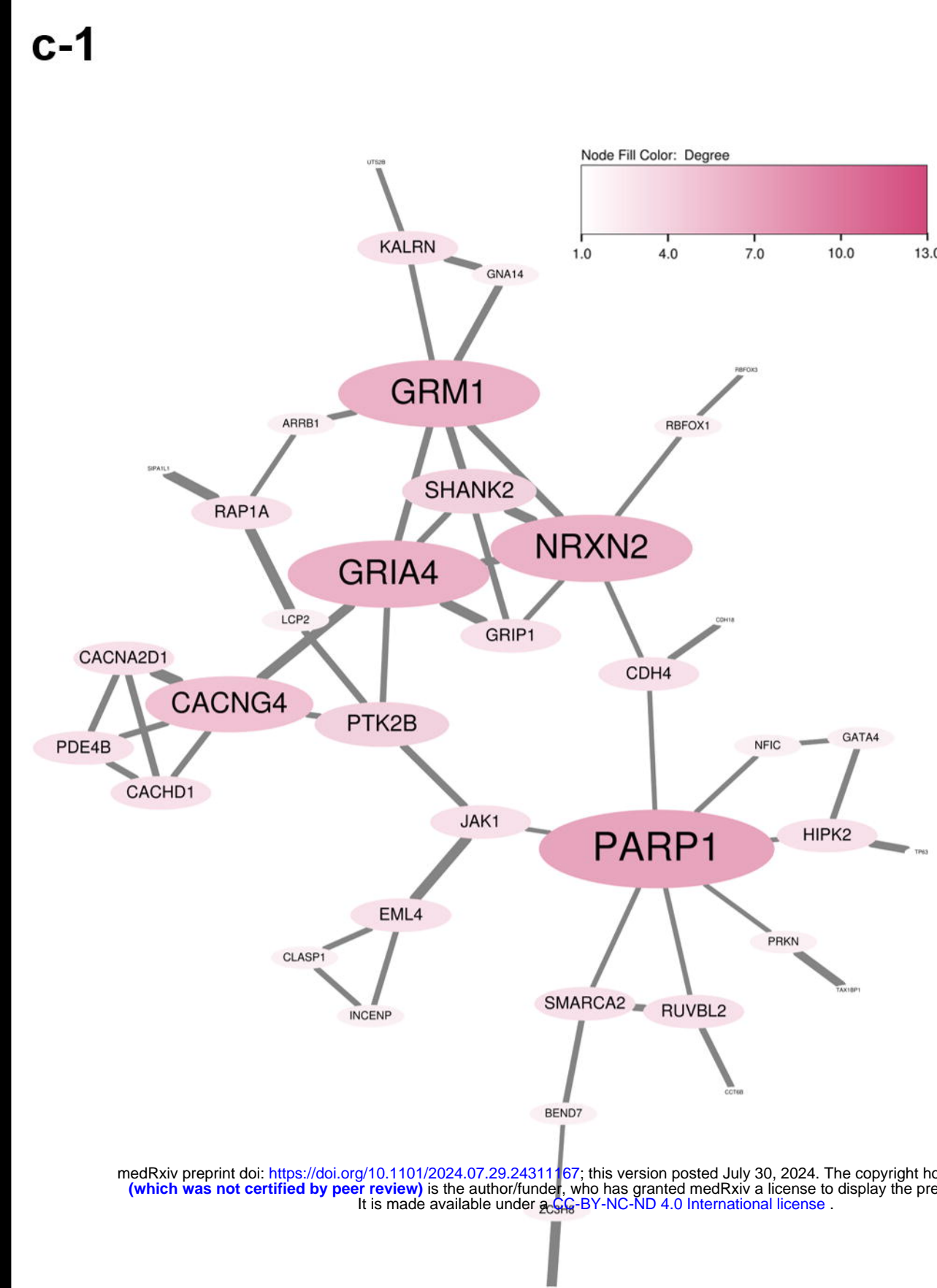
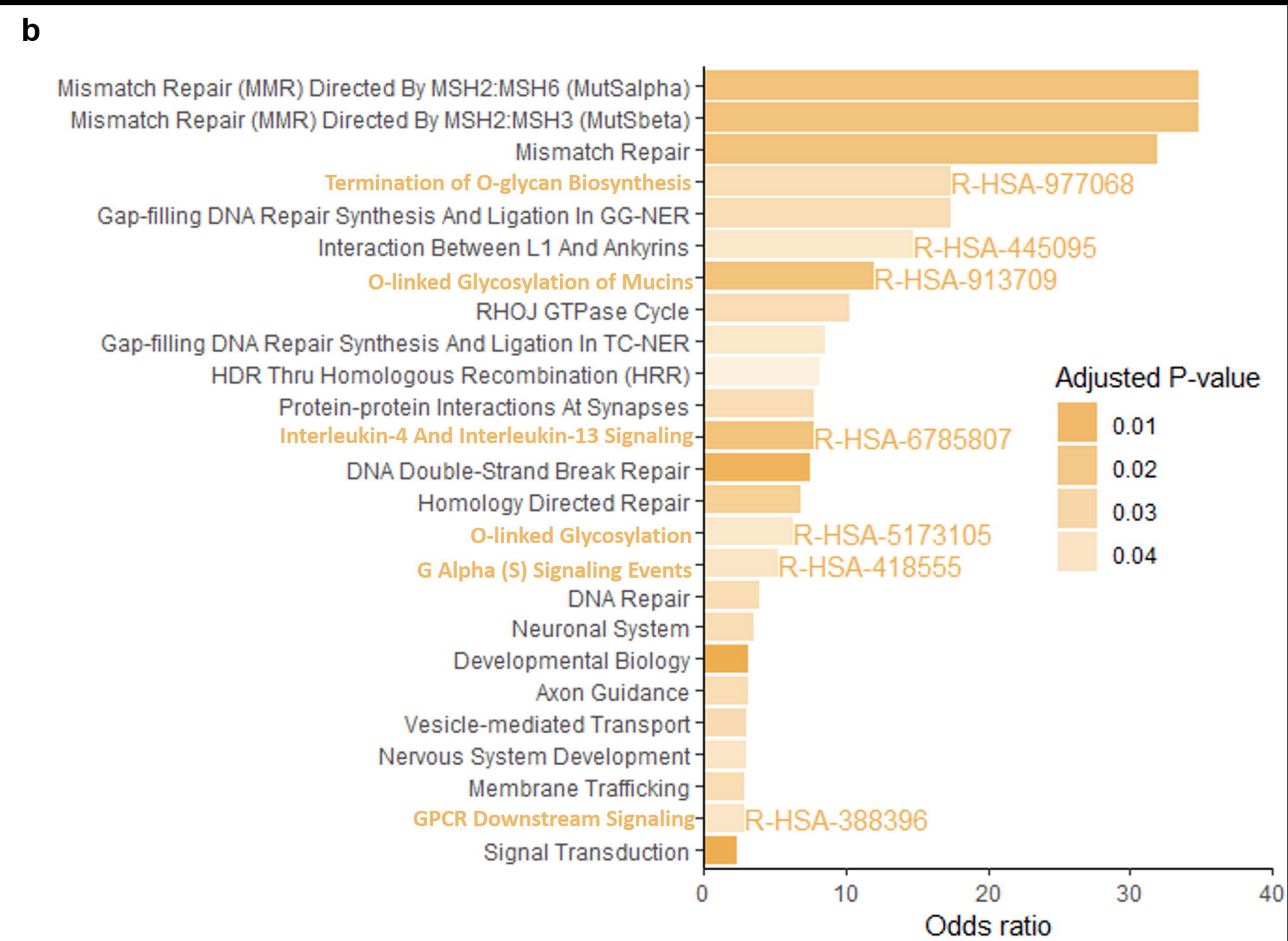
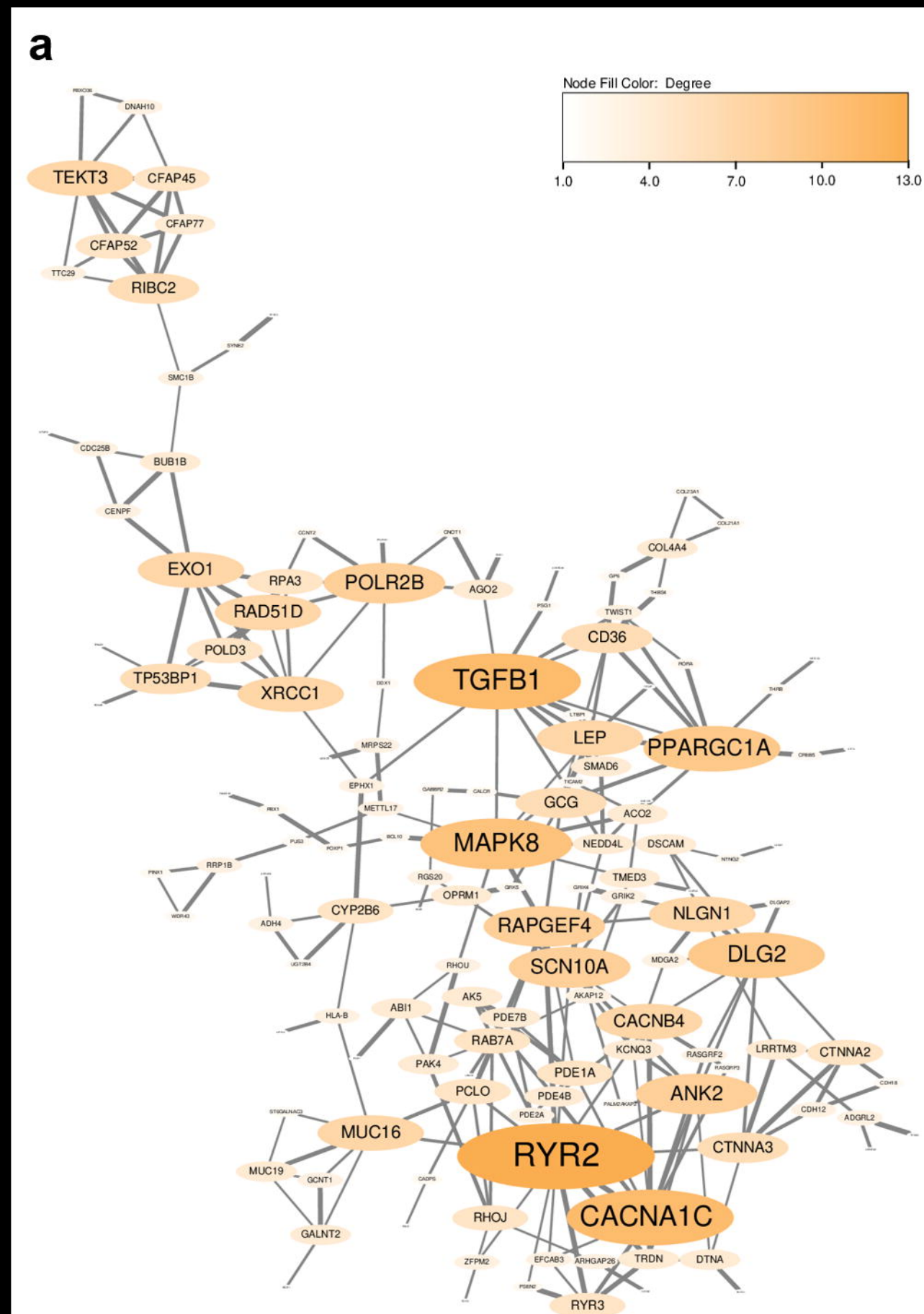
It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).



Valvular heart disease



a**b**



medRxiv preprint doi: <https://doi.org/10.1101/2024.07.29.24311167>; this version posted July 30, 2024. 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. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

