

1 **Polygenic adaptation is not a major driver of disparities in**
2 **disease mortality across global populations**

3
4 Ujani Hazra¹ and Joseph Lachance^{1*}

5
6 ¹ School of Biological Sciences, Georgia Institute of Technology, Atlanta Georgia, USA

7 * Corresponding author:

8
9 Email addresses of authors:

10 ujani.hazra@gatech.edu

11 joseph.lachance@biology.gatech.edu

12
13 **Keywords:** evolutionary genetics, health disparities, natural selection, polygenic risk
14 scores, polygenic adaptation

15
16 **Running title:** polygenic adaptation and health disparities

17 **Abstract**

18

19 **Background and objectives:** Health disparities are due to a range of socioeconomic
20 and biological causes, and many common diseases have a genetic basis. Divergent
21 evolutionary histories cause allele frequencies at disease-associated loci to differ across
22 global populations. To what extent are differences in disease risks due to natural
23 selection?

24

25 **Methodology:** Examining a panel of nine global populations, we identified which of the
26 20 most common causes of death have the largest health disparities. Polygenic risk
27 scores were computed and compared for 11 common diseases for the same nine
28 populations. We then used *PolyGraph* to test whether differences in disease risk can be
29 attributed to polygenic adaptation. Finally, we compared human development index
30 statistics and polygenic risk scores to mortality rates for each population.

31

32 **Results:** Among common causes of death, HIV/AIDS and tuberculosis exhibited the
33 greatest disparities in mortality rates. Focusing on common polygenic diseases, we
34 found that genetic predictions of disease risk varied across global populations (including
35 elevated risks of lung cancer in Europeans). However, polygenic adaptation tests
36 largely yielded negative results when applied to common diseases. Our analyses
37 revealed that natural selection was not a major cause of differences in disease risks
38 across global populations. We also found that correlations between mortality rates and
39 human development index statistics were stronger than correlations between mortality
40 rates and polygenic predictions of disease risks.

41

42 **Conclusions and implications:** Although evolutionary history contributes to
43 differences in disease risks, health disparities are largely due to socioeconomic and
44 other environmental factors.

45 **Background and Objectives**

46 Epidemiological analyses have reported major disparities in disease prevalence,
47 incidence, and mortality rates across global populations and ethnic groups (1-3).
48 Environment and lifestyle changes have been known to drive disease risks across
49 populations since the Neolithic revolution when societies shifted from foraging to
50 cultivating food. Increases in population size and changes in eating habits following the
51 agricultural revolution led to an increase in nutritional and infectious diseases and a
52 decline in the overall health of populations (4). While mortality rates from infectious
53 diseases have decreased significantly in the 20th century (5), the “transition to
54 modernity” (6) now puts the global population at a greater risk of non-communicable
55 diseases. A recent study of Sub-Saharan African populations has shown that the
56 leading causes of death have changed between 1990 and 2010 from communicable
57 diseases in children to non-communicable diseases in adults, with stroke, depression,
58 diabetes, and ischemic heart disease dominating among upper-middle-income countries
59 (7). However, the increased prevalence and mortality rates of non-communicable
60 diseases are not uniform worldwide. Disease burdens of stroke and ischemic heart
61 disease vary greatly across global populations (8). Other studies have reported that
62 men of African descent suffer the highest mortality from prostate cancer globally,
63 indicating geographic and ethnic disparities (9).

64
65 Disparities in disease prevalence and mortality rates are due to a complex combination
66 of socioeconomic, demographic, environmental, and genetic causes. Socioeconomic
67 factors like poverty and lack of access to quality treatment have increased chronic
68 kidney disease risks, in addition, to an already established racial and ethnic disparities
69 (10). Similarly, environmental factors like exposure to abandoned uranium mines have
70 increased risks of hypertension, kidney disease, and cancer in some Native American
71 populations (11). It should be noted that such disease-causing factors are often
72 interdependent and act synergistically to influence disease risk. A population’s genetic
73 makeup can also make them susceptible to certain diseases. For example, people of

74 Ashkenazi descent have a higher frequency of the *BRCA1/2* mutations, subjecting them
75 to higher risks of breast cancer ([12](#)).

76
77 Unlike monogenic diseases like sickle cell anemia or cystic fibrosis, most high mortality
78 non-communicable diseases (e.g., stroke, chronic kidney disease, and diabetes) are
79 polygenic in nature, i.e., the risks of many complex diseases are due to multiple genetic
80 variations or single nucleotide polymorphisms (SNPs). Evolutionary history, including
81 natural selection and founder effects ([13-15](#)), can cause allele frequencies of disease-
82 associated SNPs to vary among populations. These allele frequency differences can, in
83 turn, cause hereditary disease risks to vary across the globe. With the advent of
84 genome-wide association studies (GWAS), large numbers of disease-associated SNPs
85 are now known ([16](#)), and these SNPs can be used to generate polygenic predictions of
86 disease risk ([17](#), [18](#)). Importantly, shifts in allele frequencies at trait-associated SNPs
87 can be used to detect natural selection acting on polygenic diseases ([19](#)).

88
89 In this paper, we integrate country-level epidemiological data, genetic data from nine
90 global populations, and GWAS findings to address the following questions: Which
91 common causes of death exhibit the greatest disparities in mortality rates? To what
92 extent do polygenic predictions of disease risk vary across global populations? Has
93 natural selection been a major cause of differences in the risks of common hereditary
94 diseases? Are disparities in mortality rates better attributed to socioeconomic or genetic
95 factors?

96
97

98 **Methodology**

99 ***Public Health Data: Disparities in Development and Mortality Rate***

100 We collated the top 20 global causes of death (Table 1) and mortality rate data from the
101 World Health Organization (WHO) report for 2019 ([20](#)), filtering and removing the non-
102 diseased traits like road injury, falls, and self-harm. Of the 17 disease traits, we
103 removed neonatal conditions, infectious diseases (diarrhea, tuberculosis, lower
104 respiratory infection, and HIV/AIDS), cirrhosis of the liver, and stomach cancer due to

105 the lack of large numbers of publicly available GWAS SNPs. This left us with a final list
106 of 10 diseases out of the 20 most common global causes of death. Diabetes mellitus
107 was split into type 1 and type 2 diabetes, to accommodate different GWAS for each type
108 of disease, resulting in a total of 11 diseases analyzed here.

109
110 Country-specific mortality rates vary for the top 20 causes of death reported by WHO.
111 To study the disparities in these mortality rates, we focused on nine populations from
112 the 1000 Genomes Project (1KGP), which have well-defined country-level disease
113 mortality data and cover each of the major continental super populations. The nine
114 representative countries and populations in this study are the United Kingdom (GBR),
115 Spain (IBS), Nigeria (YRI), Sierra Leone (MSL), Bangladesh (BEB), Sri Lanka (STU),
116 China (CHB), Japan (JPT), and Peru (PEL).

117
118 To quantify the extent of health disparities, coefficients of variation (CV) of age-
119 standardized mortality rates were calculated among countries for each of the 20 most
120 common global causes of death:

$$121 \quad CV = \sigma/\mu$$

122 where σ is the standard deviation and μ is the mean of mortality rates in the nine
123 populations (Table 1). We also obtained 2019 Human Development Index (HDI)
124 statistics from the *United Nations Development Programme, Human Development*
125 *Reports* ([21](#)) to study whether a country's socioeconomic and development status play a
126 role in determining mortality from certain diseases. Linear regression methods were
127 applied to examine correlations between the HDI and mortality rates for the top 20
128 causes of death over the nine representative populations.

129 **Genetic Data: Population Admixture and Disease Associations**

130
131 Phase 3 data from the 1KGP was used as a reference for building admixture graphs
132 and calculating polygenic risk scores. To study polygenic adaptation, publicly available
133 summary statistics of 11 genome-wide association studies ([22-32](#)) were collated (Table
134 2), with diabetes mellitus having representative GWAS for type 1 and type 2 diabetes.
135 We only considered GWAS SNPs with a p-value $< 5 \times 10^{-8}$ (Table 2). *LiftOver* ([33](#)) was

136 used to convert the coordinates of all GWAS SNPs to the hg19 build for integration with
137 the Phase 3 data from the 1KGP.

138

139 ***Predicting differences in disease risks***

140 Predicted polygenic disease risks were calculated via polygenic risk scores (PRS). Risk
141 scores were calculated for the 11 GWAS statistics in the nine representative
142 populations using the Phase 3 data from the 1KGP. As per standard practice (18), PRS
143 for each individual were calculated as:

$$PRS = \sum_i N_i \beta_i$$

144 where N_i is an individual's dosage of the effect allele ($N = 0, 1$ or 2) of the i^{th} SNP and β_i
145 is the effect size of each SNP. PRS was then converted to a standardized normal (i.e.,
146 z-score) scale, with a global mean of zero and standard deviation of one.

147

148 Differences in predicted polygenic disease risk distribution between European (GBR
149 and IBS) and non-European (BEB, STU, CHB, JPT, YRI, MSL) populations were
150 compared using Mann-Whitney U-tests. The Mann-Whitney U-test assumes that the two
151 groups being compared (European vs. non-European in this case) are independent.
152 Since PEL is an admixed population, we exclude it from the Mann-Whitney U-Test to
153 keep the two groups independent. Correlations of mortality rates with predicted
154 polygenic disease risk scores were studied for the 11 diseases over the nine
155 representative populations.

156

157 ***Examining Polygenic Adaptation***

158 To investigate signals of polygenic adaptation, we used *PolyGraph* (34), A Markov
159 Chain Monte Carlo (MCMC) algorithm that utilizes admixture graph information to
160 deduce traces of polygenic adaptation in populations. To detect selection on a trait
161 *PolyGraph* requires a set of summary statistics from GWAS, neutral control SNPs that
162 are not associated with the trait, and an admixture graph of the representative
163 populations. Since *PolyGraph* estimates adaptation based on ancestral and derived
164 states of SNPs, it is required to have the knowledge of ancestral alleles of all the GWAS

165 hits and polarize the effect sizes of SNPs based on the derived allele. Thus, only GWAS
166 hits for which ancestral allele information was available from the 1KGP dataset were
167 used in our study. Odds-ratios (OR) reported by some GWASs were converted to effect
168 sizes, where $\beta = \ln(\text{OR})$.

169
170 Next, we used *SNPsnap* (35) to obtain frequency-matched neutral SNPs for the GWAS
171 hits. We frequency matched the GWAS SNPs with European, East Asian, and West
172 African populations from the 1KGP, and pooled from these three sets to make a
173 universal set of neutral/control SNPs for each trait. The same set of neutral SNPs was
174 used to build an admixture graph using *MixMapper* (36). We made scaffold trees with
175 eight continental populations and putatively added the population from Peru (PEL) as an
176 admixed population (note that the one branch leading to PEL represents Native
177 American ancestry). For each disease, *PolyGraph* reports a positive selection
178 parameter alpha, which is a product of the selection coefficient for the advantageous
179 allele and the duration of the selective process (34), and a p-value for selection on the
180 entire admixture graph. To correct for multiple testing, we calculated FDR-adjusted q-
181 values from the overall p-values of selection from *PolyGraph*.

182
183

184 **Results**

185 ***Disparities in Mortality Rates and Development Index***

186 The top 20 causes of death as of 2019 show considerable variation in mortality rates
187 worldwide (Table 1). The largest disparities were observed for lower respiratory
188 infections (CV = 1.05), neonatal disease (CV = 1.177), diarrhea (CV = 1.575),
189 tuberculosis (CV = 1.638), and HIV/AIDS (CV = 1.784), with the African countries
190 (Nigeria and Sierra Leone) having higher mortality rates than Europe and East Asia.
191 According to the WHO report for 2019, Japan has a negligible mortality rate for
192 HIV/AIDS compared to 32.4 deaths per age-standardized 100,000 people in Sierra
193 Leone. Stroke and ischemic heart disease cause the highest deaths in all the nine
194 populations, with the United Kingdom and Spain having the lowest mortality rates

195 among the nine countries. Mortality rates from self-harm have the lowest variation
196 among the nine countries studied here.

197

198 ***Differences in Predicted Polygenic Disease Risk***

199 Genetic prediction of disease risk varies across the nine populations (Fig. 1) for all 11
200 diseases. The predicted polygenic disease risk distributions are standardized, with a
201 global mean (z-score = 0). Here, positive z-scores indicate predicted polygenic disease
202 risks higher than the mean global predicted risk, and negative z-scores indicate risks
203 lower than the mean (Table S1 lists the mean z-score of risk of each of the nine
204 populations for the 11 diseases). Predicted polygenic disease risks are similar for
205 populations within the same continent but differ among continental groups. For
206 example, predicted polygenic risks for chronic kidney diseases vary by continent:
207 African, East Asian, and American populations have higher predicted risks than South
208 Asians and Europeans. Polygenic disease risk for lung cancer shows a wide distribution
209 in European populations, with quite a few individuals with very high predicted risk
210 scores, along with a high mean predicted risk for GBR and IBS populations. Europeans
211 exhibit a higher predicted polygenic disease risk for lung cancer on a continental scale
212 than non-Europeans (p-value = 3.43×10^{-14} , Mann-Whitney U Test). Similarly, ischemic
213 heart disease, colorectal cancer, and Alzheimer's disease also show higher predicted
214 polygenic disease risks in Europeans than non-Europeans (p-values for the Mann-
215 Whitney U Test of each disease are available in Table S2). Chronic obstructive
216 pulmonary disease narrowly passes the significance threshold to show higher risks in
217 Europe with a p-value of 0.048 (Mann-Whitney U Test). Breast cancer, stroke, and
218 chronic kidney disease have similar predicted polygenic risk score distributions in
219 European, South Asian, and East Asian populations (Table S1). Europeans have lower
220 predicted polygenic risks for type 1 diabetes (p-value = 5.3×10^{-3} , Mann-Whitney U
221 Test) and type 2 diabetes (p-value = 2.74×10^{-9} , Mann-Whitney U Test) than all non-
222 Europeans, combined. Of the diseases studied here, only hypertensive heart disease
223 does not show significant differences in predicted polygenic disease risks between
224 Europeans and non-Europeans (p-value = 0.556, Mann-Whitney U Test). Overall, we

225 see considerable variation in genetic predictions of disease risks within individual
226 populations and among different continental groups.

227

228 ***Evidence for Polygenic Adaptation***

229 *MixMapper* was used to generate the admixture graph for the nine populations (Fig. 2).
230 The graph indicated that the 1KGP population from Peru (PEL) is admixed. The graph
231 topology showed an early split of African and non-African populations, with the non-
232 Africans further branching into European and Asian groups. According to the admixture
233 graph produced by *MixMapper*, the Asian branch further divides into East and South
234 Asian sub-branches.

235

236 The strength of selection in each branch of the admixture graph was obtained from
237 *PolyGraph* for all 11 disease traits (Fig. 2). *PolyGraph* detects adaptation of polygenic
238 traits due to shifts in allele frequency of multiple loci using an admixture graph
239 framework that considers historical divergence of populations. Although *PolyGraph*
240 identifies weak signals of polygenic adaptation in all 11 traits, the FDR adjusted q-
241 values do not pass the threshold of significance (q-value < 0.05) for any of the diseases
242 studied here. We note that the raw p-value for ischemic heart disease is 0.01 over the
243 entire admixture graph, but that it does not reach statistical significance once we correct
244 for multiple testing (FDR-adjusted q-value = 0.11). *Polygraph's* overall p-values and the
245 FDR adjusted q-values for each disease are listed in Table 2.

246

247 While overall signals of polygenic adaptation are weak, certain branches of the
248 admixture graph show detectable signals of adaptation for some diseases (branch-
249 specific statistics from *PolyGraph* for each disease are listed in Table S3). Of the two
250 major heart diseases we analyzed, both ischemic heart disease and hypertensive heart
251 disease show a negative signal of adaptation for branches that ultimately lead to PEL.
252 The branches leading to GBR and IBS for lung cancer show positive selection signals,
253 consistent with the increased predicted polygenic risk for lung cancer in Europeans.
254 Adaptation signals for breast cancer are negligible in all the branches of the admixture
255 graph. Similarly, type 1 diabetes is also found to have evolved neutrally, although

256 predictive risks show Europeans are at a slightly higher risk of getting the disease.
257 Thus, although *PolyGraph* reports weak branch-specific selection signals for individual
258 diseases, the overall pattern is that polygenic adaptation is not a driver of disease
259 disparities among populations.

260

261 ***Drivers of Disparity in Mortality Rates***

262 Disparities in disease mortality rates arise from multiple socioeconomic and genetic
263 factors. To test how socioeconomic status influences mortality rates, correlations of HDI
264 to mortality rates were calculated. The mortality rates of the top 20 causes of death in
265 the nine representative populations can mostly be explained by the countries' HDI (Fig.
266 3A). Mortality rates of colorectal cancer have the highest positive correlation to HDI ($r =$
267 0.74), while mortality rates of road injuries have the highest negative correlation to HDI
268 ($r = -0.957$). The negative correlation of road injury with HDI should be interpreted with
269 caution, keeping in mind that mortality rates do not reflect incidence rates.
270 Underdeveloped countries can have higher mortality rates from road injuries due to the
271 lack of emergency medical treatment in general, despite having a lower incidence of
272 road injuries than developed countries. Alzheimer's disease, lung cancer, and stomach
273 cancer show a positive correlation with HDI, where highly developed countries have
274 higher mortality rates from these diseases (Table S4). Apart from these four, all other
275 causes of death show a negative correlation with HDI, indicating countries with lower
276 development index have higher mortality rates. An interesting observation is that
277 mortality rates of self-harm are largely independent of HDI. Highly negative correlations
278 of mortality and HDI for diseases like tuberculosis ($r = -0.781$) and HIV ($r = -0.842$) are
279 consistent with the knowledge that under-developed and developing countries bear a
280 higher burden of mortality from infectious diseases. Similarly, neonatal diseases ($r = -$
281 0.935) and diarrhea ($r = -0.914$) are major causes of infant death in under-developed
282 countries and are consistent with correlations with HDI. Thus, socioeconomic factors
283 alone play a major role in determining mortality disparities among populations.

284

285 We also examine the correlations of predicted polygenic disease risks and mortality
286 rates (Fig. 3B). Overall, the magnitudes of these correlations are lower than that of HDI

287 and mortality rates. Eight out of the ten diseases, including COPD, breast cancer,
288 diabetes, chronic kidney disease, and hypertensive heart disease, show positive
289 correlations between predicted polygenic disease risks and mortality rates, although
290 they are not statistically significant. Stroke and ischemic heart disease are interesting
291 cases with negative correlations of predicted polygenic disease risk and mortality rates.
292 Thus, although eight diseases show a weak genetic contribution to disparities in
293 mortality, genetic predictions are inconsistent with epidemiological trends when it comes
294 to stroke and ischemic heart disease. Overall, our results indicate that both
295 socioeconomic and genetic factors play a role in determining country-level mortality
296 rates.

297

298

299 **Discussion**

300 Drivers of disparities in disease incidence and mortality rates worldwide are
301 multifactorial with socioeconomic, demographic, environmental, and genetic aspects
302 (37). Contributions of genetics to disease disparities are driven by a complex
303 combination of natural selection, genetic drift, founder effect, and more (38). Here, we
304 demonstrate the role of polygenic adaptation in shaping disparities in disease mortality
305 worldwide, and the study has provided four key results. First, in analyzing the top 20
306 causes of death, mortality rates present considerable differences among countries.
307 Mortalities from infectious diseases like diarrhea and tuberculosis vary greatly between
308 continents. On the other hand, self-harm exhibited the lowest variation in mortality rates
309 worldwide. Second, predicted polygenic risk scores for the 11 common diseases depict
310 significant differences between European and non-European populations. The
311 distributions of predicted polygenic risks are consistent within continents but differ
312 among continental populations. Third, although there exist weak branch-specific
313 signatures of selection for some diseases, the overall pattern is that polygenic
314 adaptation is not a major driver of differences in disease risk across populations. And
315 finally, we show that disparities in mortality rates are highly correlated with a country's
316 HDI, indicating socioeconomic and environmental factors are strong drivers of
317 disparities.

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

Predicted polygenic disease risks of complex diseases like diabetes and lung cancer often vary among populations (Fig. 1) (39). This arises from differences in allele frequencies of disease-associated loci in different populations. For example, the C allele (rs7901695) at the *TCF7L2* locus for type 2 diabetes has an almost 40-fold increase in its allele frequency from 0.013 in the Chinese populations (CHB) to 0.488 in the African population (13), which contributes to the general pattern of higher predicted risks of type 2 diabetes in African than in East Asian populations (Fig. 1). Although we focus on differences between populations, considerable variation in predicted polygenic disease risk exists within each population, most notable in the wide distribution of predicted risks for lung cancer in Europeans (GBR and IBS) (Fig. 1). Consequently, while polygenic risk prediction has been shown to identify individuals at high risk of a disease (40), caution should be taken when trying to extend these results to broader populations (this is known as the “Ecological Fallacy” of inferring individual data from an aggregate (41)). Nevertheless, polygenic predictions of disease risks tend to be positively correlated with mortality rates (Fig. 3B).

Differences in polygenic disease risk can be due to multiple evolutionary forces, polygenic adaptation being one of them. While our results suggest weak selection for certain diseases in specific evolutionary branches, the overall lack of statistical significance leads us to rule out polygenic adaptation as a significant evolutionary cause of disease disparities. Selection of polygenic traits is distributed across multiple loci. While some loci have a greater influence on the trait with strong selection acting on them, the total selection of a polygenic trait is often mitigated by the alleles with weaker selection, making it difficult to identify overall signals of selection. This contrasts with some well-known Mendelian diseases like sickle cell anemia, where strong selection on a single locus has increased its prevalence in the African populations due to natural selection on genetic resistance to malaria (42).

Although we obtain compelling results on the drivers of disease disparity, we are aware of certain caveats in our methods. First, we study differences in mortality rates in

349 populations instead of disparities in incidence rates. The reader should be mindful that
350 while countries have disparities in mortality rates from diseases, there might not be
351 equivalent disparities in incidence rates. Second, for all the common disease studies,
352 well-powered GWASs across multiple populations were not available, and thus we
353 focused on GWASs performed in European populations. Finally, the choice of our nine
354 representative populations is highly dependent on the availability of genetic data from
355 these populations. As more GWAS are performed in varied populations worldwide,
356 future studies will be able to take a more detailed look at polygenic adaptation patterns
357 and genetic influences of disease disparities.

358

359 **Conclusion and Implications:**

360 Disparities in disease mortality rates worldwide arise from multiple factors, and here we
361 show that natural selection and polygenic adaptation are not primary contributors to
362 such disparities. We expect to form a more refined idea of how genetics and polygenic
363 adaptation influence disease disparities as newer data sets become available in the
364 future. But until then, socioeconomic and developmental factors should be thoroughly
365 studied to understand the roots of such disparities. To conclude, even though
366 populations have experienced unique evolutionary histories and many common causes
367 of diseases are heritable, our results indicate that most health disparities are driven by
368 factors other than polygenic adaptation.

369

370

371 **Funding and Conflict of Interest:**

372 This work was supported by an NIH MIRA grant to Joseph Lachance (R35GM133727).
373 The funders did not have any role in the design, analysis or writing of this article. We
374 declare no conflict of interests for both the authors.

375

376

377 **Acknowledgments**

378 We thank Aaron Pfennig, Mimi Brown, and Shashwat Nagar for their helpful comments
379 and suggestions during the writing of this manuscript.

References

- 380
381 1. Crimmins E, Kim JK, Vasunilashorn S; Biodemography: new approaches to
382 understanding trends and differences in population health and mortality.
383 *Demography* 2010;**47 Suppl**:S41-64. doi: 10.1353/dem.2010.0005.
- 384 2. Mullins CD, Blatt L, Gbarayor CM, et al.; Health disparities: a barrier to high-quality
385 care. *Am J Health Syst Pharm* 2005;**62**(18):1873-82. doi: 10.2146/ajhp050064.
- 386 3. Phelan JC, Link BG; Controlling disease and creating disparities: a fundamental
387 cause perspective. *J Gerontol B Psychol Sci Soc Sci* 2005;**60 Spec No 2**:27-33. doi:
388 10.1093/geronb/60.special_issue_2.s27.
- 389 4. Mummert A, Esche E, Robinson J, et al.; Stature and robusticity during the
390 agricultural transition: evidence from the bioarchaeological record. *Econ Hum Biol*
391 2011;**9**(3):284-301. doi: 10.1016/j.ehb.2011.03.004.
- 392 5. Armstrong GL, Conn LA, Pinner RW; Trends in infectious disease mortality in the
393 United States during the 20th century. *JAMA* 1999;**281**(1):61-6. doi:
394 10.1001/jama.281.1.61.
- 395 6. Corbett S, Courtiol A, Lummaa V, et al.; The transition to modernity and chronic
396 disease: mismatch and natural selection. *Nat Rev Genet* 2018;**19**(7):419-430. doi:
397 10.1038/s41576-018-0012-3.
- 398 7. Bigna JJ, Noubiap JJ; The rising burden of non-communicable diseases in sub-
399 Saharan Africa. *Lancet Glob Health* 2019;**7**(10):e1295-e1296. doi: 10.1016/S2214-
400 109X(19)30370-5.
- 401 8. Kim AS, Johnston SC; Global variation in the relative burden of stroke and ischemic
402 heart disease. *Circulation* 2011;**124**(3):314-23. doi:
403 10.1161/CIRCULATIONAHA.111.018820.
- 404 9. Rebbeck TR; Prostate Cancer Genetics: Variation by Race, Ethnicity, and
405 Geography. *Semin Radiat Oncol* 2017;**27**(1):3-10. doi:
406 10.1016/j.semradonc.2016.08.002.
- 407 10. Nicholas SB, Kalantar-Zadeh K, Norris KC; Socioeconomic disparities in chronic
408 kidney disease. *Adv Chronic Kidney Dis* 2015;**22**(1):6-15. doi:
409 10.1053/j.ackd.2014.07.002.

- 410 11. Lewis J, Hoover J, MacKenzie D; Mining and Environmental Health Disparities in
411 Native American Communities. *Curr Environ Health Rep* 2017;**4**(2):130-141. doi:
412 10.1007/s40572-017-0140-5.
- 413 12. Struewing JP, Hartge P, Wacholder S, et al.; The risk of cancer associated with
414 specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N Engl J Med*
415 1997;**336**(20):1401-8. doi: 10.1056/NEJM199705153362001.
- 416 13. Adeyemo A, Rotimi C; Genetic variants associated with complex human diseases
417 show wide variation across multiple populations. *Public Health Genomics*
418 2010;**13**(2):72-9. doi: 10.1159/000218711.
- 419 14. Gibson G; Hints of hidden heritability in GWAS. *Nat Genet* 2010;**42**(7):558-60. doi:
420 10.1038/ng0710-558.
- 421 15. Shifman S, Kuypers J, Kokoris M, et al.; Linkage disequilibrium patterns of the
422 human genome across populations. *Hum Mol Genet* 2003;**12**(7):771-6. doi:
423 10.1093/hmg/ddg088.
- 424 16. McCarthy MI, Abecasis GR, Cardon LR, et al.; Genome-wide association studies for
425 complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*
426 2008;**9**(5):356-69. doi: 10.1038/nrg2344.
- 427 17. Chande AT, Wang L, Rishishwar L, et al.; GlobAI Distribution of GEnetic Traits
428 (GADGET) web server: polygenic trait scores worldwide. *Nucleic Acids Res*
429 2018;**46**(W1):W121-W126. doi: 10.1093/nar/gky415.
- 430 18. Lewis CM, Vassos E; Polygenic risk scores: from research tools to clinical
431 instruments. *Genome Med* 2020;**12**(1):44. doi: 10.1186/s13073-020-00742-5.
- 432 19. Berg JJ, Coop G; A population genetic signal of polygenic adaptation. *PLoS Genet*
433 2014;**10**(8):e1004412. doi: 10.1371/journal.pgen.1004412.
- 434 20. World Health Organisation. *Global Health Estimates 2020: Deaths by Cause, Age,*
435 *Sex, by Country and by Region, 2000-2019.*
436 [https://www.who.int/data/gho/data/themes/mortality-and-global-health-](https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death)
437 [estimates/ghe-leading-causes-of-death](https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death) (Date Accessed 2020 Accessed, date last
438 accessed)
- 439 21. United Nations Development Programme. *Human Development Report 2020. The*
440 *Next Frontier: Human Development and the Anthropocene.*

- 441 <http://hdr.undp.org/en/content/human-development-report-2020> (Date Accessed
442 2020 Accessed, date last accessed)
- 443 22. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al.; Meta-analysis of 74,046
444 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*
445 2013;**45**(12):1452-8. doi: 10.1038/ng.2802.
- 446 23. McKay JD, Hung RJ, Han Y, et al.; Large-scale association analysis identifies new
447 lung cancer susceptibility loci and heterogeneity in genetic susceptibility across
448 histological subtypes. *Nat Genet* 2017;**49**(7):1126-1132. doi: 10.1038/ng.3892.
- 449 24. Malik R, Chauhan G, Traylor M, et al.; Multiancestry genome-wide association study
450 of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes.
451 *Nat Genet* 2018;**50**(4):524-537. doi: 10.1038/s41588-018-0058-3.
- 452 25. van der Harst P, Verweij N; Identification of 64 Novel Genetic Loci Provides an
453 Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ Res*
454 2018;**122**(3):433-443. doi: 10.1161/CIRCRESAHA.117.312086.
- 455 26. Xue A, Wu Y, Zhu Z, et al.; Genome-wide association analyses identify 143 risk
456 variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun*
457 2018;**9**(1):2941. doi: 10.1038/s41467-018-04951-w.
- 458 27. Giri A, Hellwege JN, Keaton JM, et al.; Trans-ethnic association study of blood
459 pressure determinants in over 750,000 individuals. *Nat Genet* 2019;**51**(1):51-62. doi:
460 10.1038/s41588-018-0303-9.
- 461 28. Law PJ, Timofeeva M, Fernandez-Rozadilla C, et al.; Association analyses identify
462 31 new risk loci for colorectal cancer susceptibility. *Nat Commun* 2019;**10**(1):2154.
463 doi: 10.1038/s41467-019-09775-w.
- 464 29. Mavaddat N, Michailidou K, Dennis J, et al.; Polygenic Risk Scores for Prediction of
465 Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 2019;**104**(1):21-34.
466 doi: 10.1016/j.ajhg.2018.11.002.
- 467 30. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al.; Genetic landscape of
468 chronic obstructive pulmonary disease identifies heterogeneous cell-type and
469 phenotype associations. *Nat Genet* 2019;**51**(3):494-505. doi: 10.1038/s41588-018-
470 0342-2.

- 471 31. Wuttke M, Li Y, Li M, et al.; A catalog of genetic loci associated with kidney function
472 from analyses of a million individuals. *Nat Genet* 2019;**51**(6):957-972. doi:
473 10.1038/s41588-019-0407-x.
- 474 32. Chiou J, Geusz RJ, Okino ML, et al.; Interpreting type 1 diabetes risk with genetics
475 and single-cell epigenomics. *Nature* 2021;**594**(7863):398-402. doi: 10.1038/s41586-
476 021-03552-w.
- 477 33. Hinrichs AS, Karolchik D, Baertsch R, et al.; The UCSC Genome Browser Database:
478 update 2006. *Nucleic Acids Res* 2006;**34**(Database issue):D590-8. doi:
479 10.1093/nar/gkj144.
- 480 34. Racimo F, Berg JJ, Pickrell JK; Detecting Polygenic Adaptation in Admixture
481 Graphs. *Genetics* 2018;**208**(4):1565-1584. doi: 10.1534/genetics.117.300489.
- 482 35. Pers TH, Timshel P, Hirschhorn JN; SNPsnip: a Web-based tool for identification
483 and annotation of matched SNPs. *Bioinformatics* 2015;**31**(3):418-20. doi:
484 10.1093/bioinformatics/btu655.
- 485 36. Lipson M, Loh PR, Levin A, et al.; Efficient moment-based inference of admixture
486 parameters and sources of gene flow. *Mol Biol Evol* 2013;**30**(8):1788-802. doi:
487 10.1093/molbev/mst099.
- 488 37. Adler NE, Rehkopf DH; U.S. disparities in health: descriptions, causes, and
489 mechanisms. *Annu Rev Public Health* 2008;**29**:235-52. doi:
490 10.1146/annurev.publhealth.29.020907.090852.
- 491 38. Benton ML, Abraham A, LaBella AL, et al.; The influence of evolutionary history on
492 human health and disease. *Nat Rev Genet* 2021;**22**(5):269-283. doi:
493 10.1038/s41576-020-00305-9.
- 494 39. Martin AR, Gignoux CR, Walters RK, et al.; Human Demographic History Impacts
495 Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet*
496 2020;**107**(4):788-789. doi: 10.1016/j.ajhg.2020.08.020.
- 497 40. Khera AV, Chaffin M, Aragam KG, et al.; Genome-wide polygenic scores for
498 common diseases identify individuals with risk equivalent to monogenic mutations.
499 *Nat Genet* 2018;**50**(9):1219-1224. doi: 10.1038/s41588-018-0183-z.
- 500 41. Piantadosi S, Byar DP, Green SB; The ecological fallacy. *Am J Epidemiol*
501 1988;**127**(5):893-904. doi: 10.1093/oxfordjournals.aje.a114892.

502 42. Allison AC; Protection afforded by sickle-cell trait against subtertian malarial
503 infection. *Br Med J* 1954;1(4857):290-4. doi: 10.1136/bmj.1.4857.290.

504

505

Tables

506

Table 1

Causes of death	United Kingdom	Spain	Nigeria	Sierra Leone	Bangladesh	Sri Lanka	China	Japan	Peru	CV
HIV/AIDS	0.2	0.8	26.8	34.2	0.4	0.2	1.6	0	2.6	1.784
Tuberculosis	0.2	0.2	99.1	43.6	25.9	3.5	1.9	0.5	7.1	1.638
Diarrheal diseases	0.9	0.6	79	91.1	29.6	2.5	0.7	0.5	3.3	1.575
Neonatal conditions	3	2.2	36	34.2	19	4	3.8	0.7	5.2	1.177
Lower respiratory infections	20.4	8	98.1	131.4	23	20.4	11.3	16.6	44.2	1.05
Cirrhosis of the liver	7.9	4.8	42	33.3	16.6	5	7.6	4.8	19.3	0.87
Diabetes mellitus	3.6	5	33.5	38	25.5	47.3	8.6	1.4	13.3	0.865
Road injury	2.8	3.1	27.3	41.6	16.7	18.1	15.1	2.2	13.4	0.826
Hypertensive heart disease	3	4.7	16.1	20.3	13.3	4.1	16.8	1.1	3.1	0.802
Trachea, bronchus, lung cancers	24.7	22.8	1.4	2.4	10.2	6.6	36.6	18.1	9.5	0.796
Stroke	20.2	18.4	88.7	120.1	119.2	26.6	110.8	23.2	22.6	0.77
Stomach cancer	3.3	5.2	2.5	7	5.5	4.9	20.5	11.5	15.1	0.723
Kidney diseases	2.9	5.5	23.9	30.2	9.1	26.2	10.4	7.6	21.3	0.666
Chronic obstructive pulmonary disease	19.8	12.3	20.8	27.4	41.7	37	54.2	12.7	7.2	0.603
Falls	4.3	2.5	8.5	9	2.6	3.9	8	2	3.8	0.559
Colon and rectum cancers	12.1	14.3	4.3	5.4	3.6	3.4	13	13.6	7.9	0.535
Ischemic heart disease	43.1	34.9	100.9	139.4	94.3	101.3	97.6	30.6	40	0.515
Self-harm	6.9	5.3	6.9	11.3	3.9	12.9	6.7	12.2	2.7	0.48
Alzheimer disease and other dementias	42.7	21.5	19.4	19.3	13.9	27.6	17.4	7.9	16.2	0.477
Breast cancer	9.4	6.9	10.4	14.9	5.3	5.3	4.7	5.4	6	0.445

507

508 **Table 1.** Disparities in mortality rates for the top 20 worldwide causes of death. The top
509 20 causes of death from the World Health Organization Report of 2019 are reported
510 here. Population-specific mortality rates for each disease are age-standardized per
511 100,000 individuals. The 1000 Genomes Project population codes for each country are
512 United Kingdom (GBR), Spain (IBS), Nigeria (YRI), Sierra Leone (MSL), Bangladesh
513 (BEB), Sri Lanka (STU), China (CHB), Japan (JPT) and Peru (PEL). CV stands for the
514 coefficient of variation of mortality rates. Note that mortality for type 1 diabetes and type
515 2 diabetes are reported jointly as diabetes mellitus.

516

Table 2

Disease	GWAS source	# of SNPs	Overall p-value	FDR adjusted q-value
Ischemic heart disease	van der Harst et al. (2018) (25)	148	0.01	0.11
Stroke	Malik et al. (2018) (24)	26	0.176	0.323
Chronic obstructive pulmonary disease	Sakornsakolpat et al. (2019) (30)	82	0.829	0.829
Trachea, bronchus, lung cancers	McKay et al. (2017) (23)	115	0.065	0.297
Alzheimer disease and other dementias	Lambert et al. (2013) (22)	19	0.64	0.782
Type 1 Diabetes mellitus	Chiou et al. (2021) (32)	33	0.741	0.815
Type 2 Diabetes mellitus	Xue et al. (2018) (26)	139	0.331	0.52
Kidney diseases	Wuttke et al. (2019) (31)	308	0.152	0.323
Hypertensive heart disease	Giri et al. (2019) (27)	258	0.081	0.297
Colon and rectum cancers	Law et al. (2019) (28)	40	0.175	0.323
Breast cancer	Mavaddat et al. (2019) (29)	313	0.485	0.667

517

518

519

520

521

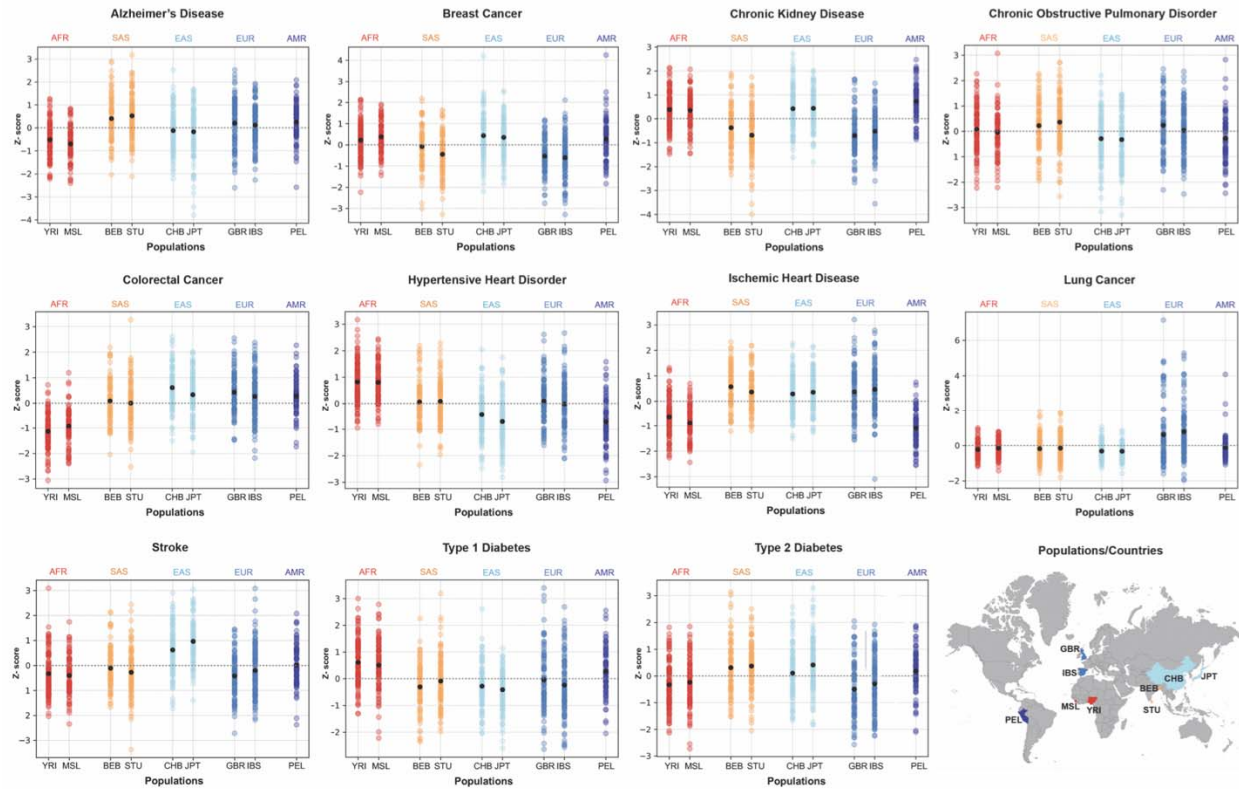
522

Table 2. Tests of polygenic adaptation for 11 common diseases. The number of disease-associated loci reported and the reference GWAS are listed for each disease. *Polygraph* reports the overall p-value as a measure of the significance of polygenic adaptation in the entire admixture graph. FDR adjusted q-values correct for multiple testing.

523 **Figures**

524

525 **Figure 1**



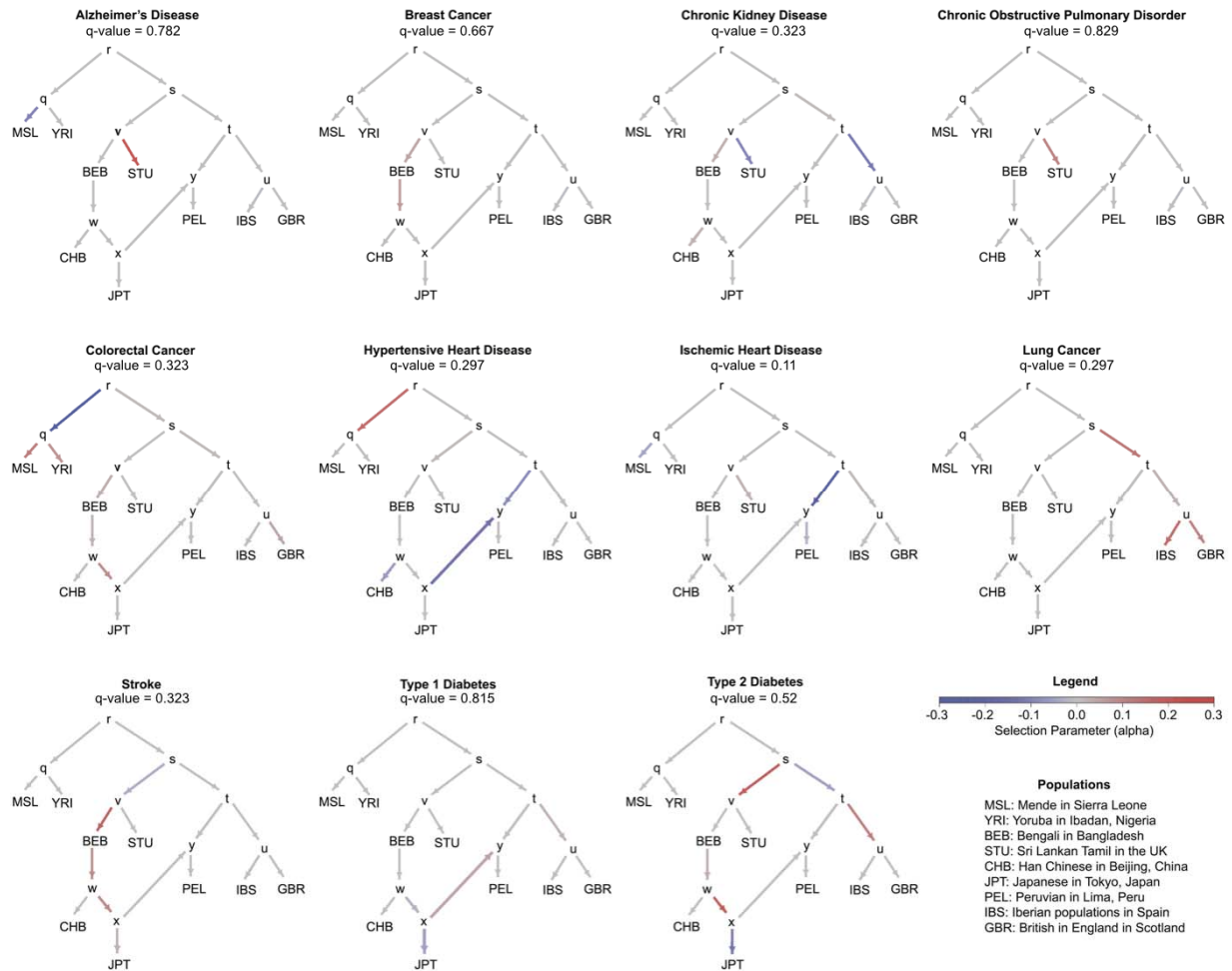
526

527

528 **Figure 1.** Predicted polygenic risks for 11 common diseases over the nine
529 representative populations. Z-score distributions of predicted polygenic disease risks
530 are shown here. Each dot in the distribution represents an individual. Distributions
531 above the global mean (z-score = 0) indicate higher risks, whereas those below zero
532 indicate lower disease risks. Different colors indicate continental superpopulations.

533

Figure 2



534

535

536

537

538

539

540

541

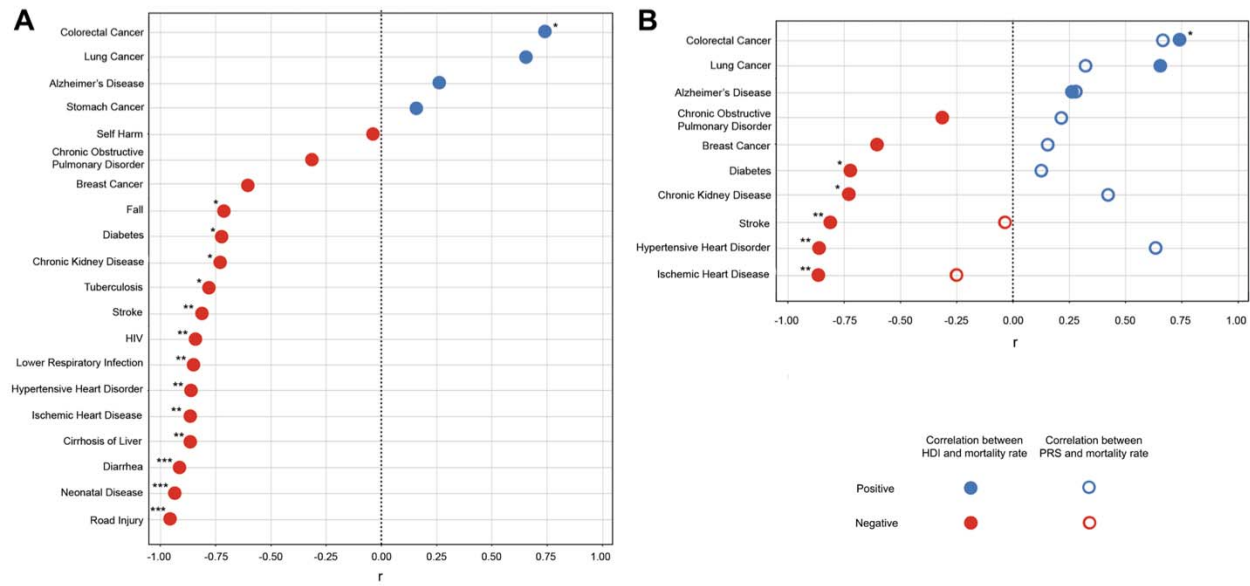
542

543

Figure 2. Minimal evidence of natural selection acting on common polygenic diseases. *MixMapper* was used to generate the admixture graph and *PolyGraph* was used to test for polygenic signatures of adaptation. FDR adjusted q-values are above 0.05 for all nine diseases. The selection parameter alpha reports a product of the selection coefficient for the advantageous allele and the duration of the selective process. Positive values of alpha indicate positive selection in the branch while negative values are indicative of negative selection acting on a trait.

544

Figure 3



545

546

547

548

549

550

551

552

553

554

Figure 3. Correlations of mortality rates with Human Development Index (HDI) statistics and predicted genetic disease risks. **(A)** 14 out of the top 20 causes of death from the 2019 World Health Organization report show a statistically significant correlation between country-specific mortality rates and HDI. **(B)** Correlations of mortality with HDI and polygenic risk scores. Filled circles indicate correlations between mortality rates and HDI, and open circles indicate correlations between mortality rates and PRS (positive correlations in blue, negative correlations in red). * indicates p-value < 0.05, ** indicates p-value < 0.01 and *** indicates p-value < 0.001.