The evolutionary genetics of polygenic diseases with the largest global burden in mortality rates

Ujani Hazra[1] and Joseph Lachance[1]*

*Corresponding author: Joseph Lachance (joseph.lachance@biology.gatech.edu)

Keywords: evolutionary genetics, health inequities, human genomics, natural selection, neutral evolution, polygenic diseases, ascertainment bias

***Corresponding author**: joseph.lachance@biology.gatech.edu

Author affiliations

[1] School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

Abstract

Background: Although the prevalence of many complex diseases varies across human populations, the extent to which differences in the risks of common polygenic diseases are due to local adaptation is largely unknown. Focusing on the ten hereditary diseases with the largest global disease burden in terms of mortality rates, we leveraged GWAS findings from multiple ascertainment schemes to quantify collective signatures of selection acting on sets of disease-associated variants. **Results:** First, we used PolyGraph to assess whether risk- associated loci have experienced directional shifts in allele frequencies. For each disease, these tests of polygenic adaptation revealed that most evolutionary branches did not exhibit any concerted change in predicted disease risks. Next, we developed a novel approach to quantify whether sets of disease-associated SNPs were enriched for outlier values from scans of selection compared to matched sets of control SNPs. Applying this approach to integrative haplotype scores, we did not observe strong signals of recent positive selection acting on common polygenic diseases. By contrast, application of our outlier approach to McVicker's B statistics revealed that disease-associated SNPs are enriched for signatures of background selection. Furthermore, these tests of negative selection yielded consistent patterns regardless of whether disease-associated SNPs were ascertained in European, East Asian, or multi-ancestry cohorts. **Conclusions:** While our results do not support a major role for recent positive selection or local adaptation in shaping population differences in polygenic disease risks, they do suggest that background selection continues to act on disease-associated loci across diverse human populations.

Introduction

Disease risks have evolved substantially over recent human history [1, 2]. Increases in population size and changes in eating habits following the agricultural revolution led to an increase in nutritional and infectious diseases and a decline in the overall health of many ancient populations [3]. More recently the transition to modernity following the Industrial Revolution has resulted in increased rates of chronic non-communicable diseases [4, 5]. Indeed, the leading causes of death in sub-Saharan Africa have shifted from communicable diseases in children to non-communicable diseases in adults over the past three decades, with stroke, depression, diabetes, and ischemic heart disease dominating among middle-income countries [6]. The unique evolutionary histories experienced by different human populations can also cause disease risks to diverge.

Substantial heterogeneity in the mortality rates of non-communicable diseases exists across the globe, and these health inequities arise from a complex combination of socioeconomic, demographic, environmental, and genetic causes [7, 8] (Fig. 1). The narrow sense heritabilities of many complex diseases exceed 30% [9], and allele frequencies at disease-associated loci contribute to differences in hereditary disease risks. These population-level differences can be due to natural selection [10] or neutral processes like genetic drift and population bottlenecks [11, 12]. For example, individuals of African descent have higher rates of sickle cell disease due to evolutionary tradeoffs and selection against malaria [13]. By contrast, higher risks of cystic fibrosis in the Saguenay-Lac St. Jean region of Québec are due to founder effects [14]. However, unlike

these two examples, many common diseases have complex non-Mendelian genetic architectures.

One key insight gleaned from two decades of genome-wide association studies (GWAS) is that most non-communicable diseases are polygenic, i.e., hereditary disease risks are due to the cumulative effects of multiple SNPs [15-17]. These genetic architectures require polygenic tests of natural selection that focus on sets of trait-associated SNPs [18-20]. Analyses of ancient DNA have painted a mixed picture of human evolution during the Holocene (due in part to data quality issues and the need to correct for population continuity): some complex traits have appear to be evolving neutrally while others show evidence of polygenic selection [21-23]. Bayesian approaches examining the relationship between minor allele frequencies and the effect sizes of disease-associated SNPs have shown that negative and/or stabilizing selection widely impacts genetic variants that are associated with complex polygenic traits [24-26]. Similarly, recent analyses of conditional frequency spectra reveal that complex traits have experienced purifying and/or stabilizing selection [27]. Disease-associated SNP sets also exhibit signatures of evolutionary conservation over longer phylogenetic timescales [28]. One major limitation of most published studies of polygenic selection is that they focus on a single SNP set for each disease (most often from a European ancestry GWAS), and as such they do not control for ascertainment bias. Nor have published studies of polygenic selection focused on diseases with the largest global health burden. An open question is whether local adaptation and recent selection can explain why the mortality rates of these

diseases vary substantially across populations or whether these differences are primarily due to socioeconomic factors (Fig. 1).

Here, we leveraged GWAS findings from European, East Asian and multi-ancestry cohorts to infer whether sets of disease-associated SNPs are enriched for signatures of natural selection (Table 1). We addressed three questions: 1) To what extent are population-level differences in hereditary disease burdens due to recent positive selection and polygenic adaptation? 2) Do complex diseases exhibit signatures of background selection? 3) Are our findings consistent across different GWAS ascertainment schemes and estimates of SNP effect sizes?

Results

Our tests of polygenic selection focused on the top 10 hereditary diseases with the largest public health burden as reported by the World Health Organization (WHO) [29]. These complex polygenic diseases included cardiometabolic diseases, (ischemic heart disease, hypertensive heart disease, stroke, type 2 diabetes), two chronic conditions (chronic obstructive pulmonary disease (COPD) and chronic kidney disease), three types of cancer (lung cancer, colon cancer, breast cancer), and Alzheimer's disease (Fig. 1). For each disease we leveraged SNP sets from well-powered European, East Asian, and multi-ancestry GWAS cohorts (Table 1 and Table S1).

Limited signatures of polygenic adaptation

First, we used PolyGraph to quantify signatures of polygenic adaption acting on sets of disease-associated SNPs. This approach leverages allele frequencies to identify directional shifts in predicted trait values along different evolutionary branches in an admixture graph [19]. FDR-adjusted q-values in Table 1 quantify whether a null hypothesis of neutral evolution can be rejected across an entire admixture graph. 20 of the 28 disease-GWAS ancestry combinations we analyzed did not show any statistically significant signature of polygenic adaptation. Focusing on SNP sets that were ascertained in European ancestry cohorts, we found that only two out of ten diseases listed in Table 1 have FDR-adjusted q-values of less than 0.05. Broadly similar patterns were observed for PolyGraph analyses that utilized SNP sets that were ascertained using East Asian or multi-ancestry GWAS cohorts (Table1). Five out of nine disease-associated SNP sets that were ascertained using East Asian GWAS had FDR-adjusted q-values less than 0.05 and only one out of nine disease-associated SNP sets that were ascertained using multi-ancestry GWAS cohort, had FDR-adjusted q-values that were less than 0.05. Moreover, most of the individual branches in our PolyGraph analyses do not exhibit statistically significant departures from neutrality (426 out of the 504 branch-specific Q_B statistics listed in Table S2 have p-values > 0.05). Overall, the disease-associated SNP sets analyzed here appear to have experienced limited amounts of polygenic adaptation during recent human history.

Despite the general patterns described in the previous paragraph, some exceptions exist. PolyGraph results for disease-GWAS ancestry combinations that have FDR-adjusted q-values < 0.05 are shown in Fig. 2. Here, red shading indicates branches where risk-increasing alleles have increased in frequency, blue shading indicates

127 branches were risk-increasing alleles have decreased in frequency, and gray shading
128 indicates branches where there has been minimal directional change in the frequences
129 of risk-increasing alleles. Despite nominal signatures of polygenic adaptation, the
130 admixture graphs shown in Fig. 2 have a preponderance of gray branches. Furthermore,
131 branch-specific signatures of polygenic adaptation for each disease appear to be
132 sensitive to ascertainment bias, i.e. it matters whether disease-associated SNP sets
133 come from European, East Asian, or multi-ancestry cohorts. For example, the branches
134 driving the low FDR-adjusted q-values for chronic kidney disease and hypertensive heart
135 disease vary for SNP sets that were ascertained using European or East Asian cohorts
136 (compare panels A vs. D and panels B vs. E in Fig. 2). Similarly, only one lung cancer
137 SNP set yields evidence of polygenic adaptation on the IBS_u branch (Table S2, p-values
138 = 0.602, 0.562 and 0.00462 for European, East Asian, and multi-ancestry SNP sets,
139 respective). Collectively, our PolyGraph analyses did not find a strong, replicable
140 signature of polygenic adaptation on the diseases of interest.
141

142 **Polygenic tests of outlier enrichment**

143 Unlike PolyGraph, most other tests of selection focus on individual SNPs or genomic
144 regions. These tests include integrative haplotype scores (iHS), which identify loci that
145 have experienced recent positive selection [30], and McVicker's B statistics, which identify
146 loci under background selection [31]. Inferring the collective signatures of natural
147 selection acting on polygenic diseases required the development of an outlier enrichment
148 approach. Our novel approach leveraged effect size estimates from GWAS with and a
149 kernel density estimation (KDE) framework to infer whether sets of disease-associated

150 SNPs were more likely to contain outlier values of selection statistics compared to
151 matched sets of control SNPs. These outlier enrichment analyses resulted in percentile
152 ranks for each combination of disease, ascertainment scheme, and selection statistic,
153 with higher percentiles indicating an excess of outlier values compared to control SNP
154 sets. Additional details of our approach can be found in the Methods section.
155

156 **Sparse signatures of recent positive selection**

157 To identify diseases under recent positive selection, we employed our outlier enrichment
158 approach to integrated Haplotype Score (iHS) statistics from the 1000 Genomes Project
159 (1KGP) [32]. Note that this approach is complementary to PolyGraph, as iHS statistics
160 quantify extended haplotype homozygosity within each population while PolyGraph relies
161 on allele frequency differences between populations. In total, 26 global populations were
162 analyzed for 10 diseases and three ascertainment schemes (apart from Alzheimer's
163 disease which lacked East Asian and multi-ancestry SNP sets).

164 Overall, we found that disease-associated SNP sets were not enriched for iHS
165 outliers when compared to control SNP sets (Fig. 3 and Table S3). Focusing on European
166 ascertained GWASs, most diseases show low percentile values in all 26 populations, i.e.
167 they were not enriched for signatures of recent positive selection. Only six out of 260 tests
168 had percentile ranks above 95 (p-value = 0.991, one-tailed binomial test with alpha =
169 0.05). Exceptions include ischemic heart disease in South Asian populations, namely
170 Bengalis in Bangladesh (BEB), Indian Telugu in the UK (ITU), and Sri Lankan Tamil in
171 the UK (STU). The admixed population from Lima, Peru (PEL) showed enrichment for
172 high iHS statistics for chronic kidney disease and hypertensive heart disorder. However,

173 when the same diseases were tested without considering effect sizes (i.e., all SNPs were
174 given equal weight in the KDE), the iHS outlier enrichments disappeared (Table S4 and
175 Fig. S1). This supports the notion that most disease-associated SNP sets are not under
176 strong positive selection, and that potential exceptions may be due to a small number of
177 lead SNPs with large effect sizes.

178 We observed similar results for disease associations that were ascertained in East
179 Asian and multi-ancestry cohorts. Only five out of 234 tests of East Asian-ascertained
180 SNP sets showed strong enrichments for recent positive selection (p-value = 0.992, one-
181 tailed binomial test with alpha = 0.05). Similarly, only five out of 234 tests of that were
182 ascertained using multi-ancestry SNP sets also showed strong enrichments for recent
183 positive selection (p-value = 0.992, one-tailed binomial test with alpha = 0.05).
184 Importantly, none of the high iHS percentile exceptions replicate across all three SNP
185 ascertainment schemes.

186

187 **Evidence of background selection**

188 Although we found only minimal signatures of polygenic adaptation and recent positive
189 selection, disease-associated loci may be subject to other types of selection, including
190 background selection (the reduction of genetic diversity at a locus due to negative
191 selection against linked deleterious alleles). With this in mind, we used McVicker's B
192 statistics [31] and our outlier enrichment approach to measure whether sets of disease-
193 associated loci were enriched for signatures of background selection. Fig. 4 shows the
194 percentile rank for each set of disease-associated SNPs compared to matched control
195 sets for three different GWAS ascertainment schemes (see Table 1 for exact values).

196 SNP sets that were ascertained in European ancestry populations yielded percentile
197 ranks that ranged from 88.0 (colon cancer) to above 99.9 (chronic kidney disease and
198 hypertensive heart disease), indicating that disease-associated SNPs are more likely to
199 have outlier values of McVicker's B statistics compared to matched controls. Eight out of
200 ten diseases had percentile ranks above the 95th percentile of matched control sets, a
201 statistically significant fraction (p-value = 1.61 x 10^{-9}, one-tailed binomial test with alpha
202 = 0.05). We see similar patterns for disease-associated SNP sets that were ascertained
203 in East Asian populations: six out of nine diseases have percentile ranks above 95 (p-
204 value = 1.15 x 10^{-6}, one-tailed binomial test with alpha = 0.05). Likewise, seven out of
205 nine diseases have percentile ranks over 95 when SNP sets that were ascertained in
206 multi-ancestry GWAS were analyzed (p-value = 2.57 x 10^{-8}, one-tailed binomial test with
207 alpha = 0.05). Overall, we found that disease-associated SNPs were strongly enriched
208 for signatures of background selection regardless of whether the disease-associated loci
209 were ascertained using a European, East Asian, or multi-ancestry cohort.

210 Stroke, COPD, chronic kidney disease, and hypertensive heart disease showed
211 the strongest enrichment for background selection (percentile rank above 99.9 for all
212 ascertainment schemes). By contrast, lung cancer SNP sets did not show any significant
213 enrichment for signatures of background selection, perhaps due to a lack of historical
214 selection pressures against smoking during most of human history [33]. To ensure the
215 robustness of our findings, we tested whether our results were consistent across different
216 outlier thresholds of McVicker's B (Table S5 and Fig. S2). Once again, SNPs associated
217 with diseases that have a large global health burden showed enrichment for signatures

of background selection, though this trend was weaker for more stringent outlier thresholds. We further tested the consistency of our background selection results by not weighting the KDE by effect size (Table S6 and Fig. S3). Focusing on European-ascertained disease associations and giving all SNPs in each SNP set equal weight, we found that nine out of ten diseases were still enriched for signatures of background selection (p-value = 1.87×10^{-11}, one-tailed binomial test with alpha = 0.05).

Discussion

Focusing on the ten diseases with the largest global health burden, we tested whether sets of disease-associated SNPs are enriched for signatures of natural selection. PolyGraph analyses and outlier enrichment tests of iHS statistics revealed minimal evidence for polygenic adaptation and recent positive selection acting on the diseases studied here. By contrast, outlier enrichment tests of McVicker's B statistics revealed that disease-associated SNPs tend to be found in genomic regions that have lower amounts of genetic diversity, even after controlling for allele frequency and distance to the nearest exon. While SNP sets that are associated with polygenic diseases show limited signatures of local adaptation, our results reinforce the role of background selection as a long-standing evolutionary influence on disease-associated loci across human lineages.

SNP ascertainment bias presents a major challenge when using GWAS data [34, 35], and the ability to detect genetic associations depends on both allele frequencies and effect sizes in the discovery population [36]. One byproduct of this is that sets of disease-associated SNPs can vary across studies, particularly when the genetic ancestries of

240 study participants differ. When possible, we analyzed three different ascertainment
241 schemes for each disease, i.e., SNP sets that were ascertained in European, East Asian,
242 and multi-ancestry cohorts (Table 1). While some weak signals of polygenic adaptation
243 were identified, these findings were not replicated across different ascertainment
244 schemes.

245 We note that the diseases studied here are highly polygenic, and although they
246 might not show strong selection signatures at the overall trait level, individual SNPs can
247 still be under appreciable selection pressures [37, 38]. This point is magnified for species
248 with small historic effective population sizes like our own, and for diseases with late ages
249 of onset [39-41]. Disease-associated alleles can also affect multiple traits. Because of
250 this, antagonistic pleiotropy could in principle dilute directional signals of selection [42].
251 Furthermore, our study focused on common variants associated with each disease, and
252 future studies will benefit from examining the impact of rare variants on the population-
253 specific genetic architectures of complex diseases [43, 44]. Additionally, future studies
254 that combine genetic and environmental data will provide deeper insights into the risks of
255 common complex diseases and why they differ across diverse human populations [45,
256 46].

257

258 Conclusions

259 One implication of our study is that polygenic risk score (PRS) differences between
260 populations should not be interpreted as prima facia evidence of local selection.
261 Differences in (predicted) hereditary disease risks across populations may well be due to

genetic drift [47, 48] and/or ascertainment bias [34]. Although our findings do not point to local adaptation as a primary driver of health inequities, they do suggest that background selection continues to act on disease-associated loci across diverse human populations.

Methods

Processing of GWAS results

To investigate the evolutionary genetics of these complex polygenic diseases, we extracted GWAS summary statistics from ten diseases across three distinct types of study cohorts: European, East Asian, and multi-ancestry (Table 1). Due to an insufficient number of significant associations identified for Alzheimer's Disease in East Asian and multi-ancestry ascertained GWAS, this trait was excluded from our ascertainment bias testing. GWAS sample sizes exceeded 72,000 individuals for each combination and disease and study cohort (Table S1). Marginally significant autosomal SNPs with a p-value $< 5x10^{-5}$ were extracted from each GWAS. Subsequently, LD clumping was performed to isolate independent associations with an $r^2 < 0.2$ within the respective ascertained population, utilizing Plink 1.9 [49] and 1KGP phase 3 data [32] as a reference. In the case of multi-ancestry GWASs, clumping was performed separately in European, East Asian, and African reference populations from the 1KGP, and an intersection of independent associations across populations were used for further analysis (Details of SNPs pre- and post-clumping can be found in Table S1. To ensure uniformity, the LiftOver tool [50] was employed to convert coordinates of all GWAS SNPs to the GRCh37 assembly (hg19 build) of the human reference genome. When these analyses were

284 conducted well-powered GWAS results for Alzheimer's disease were not available for
285 East Asian or multi-ancestry cohorts.
286
287 **Generating control sets of SNPs**
288 In all our analyses, autosomal control SNPs were obtained using SNPSnap [51]. Each
289 trait associated SNP had approximately 5000 matched independent ($r^2 < 0.2$ in respective
290 population) SNPs. Matching criteria included allele frequency (+/- 5%), number of SNPs
291 in LD ($r^2 < 0.5$) with the target SNPs (+/- 10%), gene density (+/- 10%) in the ascertained
292 population, and distance to the nearest exon (+/- 5%). This latter filter is particularly
293 important, as simply selecting control SNPs based on distance to nearest gene would
294 treat intronic SNPs the same as exonic SNPs. SNPSnap automatically removes any
295 SNPs within the HLA region when matching for control SNPs. For European and East
296 Asian ascertained GWAS, controls SNPs were matched within their respective
297 populations from the 1KGP. In the case of multi-ancestry studies, control SNPs were
298 matched across pooled data from European, East Asian, and African populations. Lists
299 of GWAS and matched sets of control SNPs are available on GitHub (see Data
300 Availability). We note that the top GWAS hits for a trait can often capture a sufficient
301 fraction of the total SNP heritability of polygenic traits, and there are diminishing returns
302 for polygenic scores that incorporate more than the top few hundred independent hits [52-
303 54].
304
305 **PolyGraph analyses (polygenic adaptation)**

306 To investigate signals of polygenic adaptation, we used PolyGraph [19], a Markov Chain
307 Monte Carlo (MCMC) algorithm that leverages admixture graph information to detect
308 traces of polygenic adaptation in populations. PolyGraph detects adaptation of polygenic
309 traits due to allele frequency shifts at multiple loci using an admixture graph framework
310 that considers the historical divergence of populations. It uses the ancestral and derived
311 allele frequencies for each disease-associated SNP for every population along with effect
312 sizes and compares them to control distributions. In contrast to iHS statistics, which are
313 ideal for detecting hard sweeps, PolyGraph analyses are able to detect soft sweeps from
314 standing genetic variation.

315 We used control SNPs from the enrichment analyses to build an admixture graph
316 with *MixMapper* v2.0 [55]. In the original PolyGraph paper, the authors caution against
317 using programs like *TreeMix* [56], which scale drift values by the heterozygosity of
318 ancestral nodes, as this could alter the interpretation of results from PolyGraph [19].
319 Although *MixMapper* and *TreeMix* give similar topologies, we used *MixMapper* to build
320 the admixture graph to ensure accurate results. We made scaffold trees with eight
321 continental populations and added the population from Peru (PEL) as an admixed
322 population. We ran PolyGraph using 1,000,000 MCMC steps. PolyGraph reported a
323 selection parameter alpha for each disease, a product of the selection coefficient for the
324 advantageous allele and the duration of the selective process, and a p-value for selection
325 on the entire admixture graph. To correct for multiple testing, we calculated FDR-adjusted
326 q-values from the overall p-values of selection from PolyGraph.

327

328 **Trait-level distributions of selection statistics**

329 To integrate SNP-level information from test statistics into a comprehensive trait-level
330 distribution, we employed kernel density estimation (KDE). This method allowed us to
331 derive a probability distribution of the test statistic for each trait. Unlike traditional
332 estimation techniques, KDE is a nonparametric approach that does not assume that the
333 data follows a known distribution. In our implementation, we opted for a Gaussian kernel
334 and conducted a five-fold cross-validation using GridSearchCV [57] to determine the
335 optimal kernel bandwidth for the KDE. Since each associated SNP also has an effect size
336 or beta, we also weighted the SNPs according to their absolute effect sizes. Thus, outliers
337 count more for disease-associated SNPs that have larger effect sizes. Finally, we
338 obtained a probability density function of the test statistic for each set of disease-
339 associated SNPs from the KDE.

340 For each set of disease-associated SNPs, we obtained 1000 sets of matched
341 control SNPs matched for allele frequency, linkage disequilibrium (LD) patterns in the
342 ascertained populations, gene density, and distance to the nearest exon. For each SNP
343 set, we identified the proportion of SNPs, weighted by effect size, that exceeded standard
344 outlier thresholds (see below). Enrichment tests involved comparing outlier proportions of
345 disease-associated SNP sets to control sets to generate a percentile rank, with higher
346 percentiles indicating greater trait-level signatures of selection (Fig. S4). Our approach
347 differed from that of other research teams [28] in that we looked for outlier enrichment, as
348 opposed to trait averages, plus we weighted each SNP by effect size.

349

350 **Outlier enrichment of iHS (recent positive selection)**

351 We used iHS to measure recent positive selection. This statistic compares the haplotype
352 homozygosity surrounding an ancestral allele to that of the derived allele at the same
353 position. iHS values are normalized, with iHS > 1.96 indicating that haplotypes with the
354 ancestral allele are longer and positively selected than those with the derived alleles,
355 while iHS < 1.96 indicates that the derived allele has longer haplotypes. Therefore, both
356 extreme positive and negative iHS values (|iHS| > 1.96) were considered indicative of
357 selection [58]. Our analyses used pre-computed values of iHS statistics in 26 global
358 populations from the 1KGP [30].

359 To detect trait-level signatures of positive selection, we used the probability density
360 function for iHD to calculate the weighted proportion of trait-associated variants with |iHS|
361 > 1.96. Here, we used area under the curve (AUC) statistics to quantify the proportion of
362 a SNP set with outlier values of iHS values. Our test of enrichment involved generating a
363 percentile rank for the trait by comparing the iHS outlier AUC for the trait-associated SNP
364 set to 1000 AUC values for matched control sets of SNPs. Our test of enrichment involved
365 generating a percentile rank for the trait by comparing outlier AUC statistics for the trait-
366 associated SNP set to 1000 matched outlier AUC statistics for matched control sets of
367 SNPs (Fig. S4). High percentile ranks signify that trait-associated SNPs are enriched for
368 extreme iHS values. We also tested whether our recent positive selection results were
369 robust to ignoring effective size differences – focusing on European ancestry GWAS hits
370 (Table S4 and Fig. S1). This involved modifying the iHS KDE by using the same effect
371 size for every SNP, i.e. by weighting each SNP equally.

372

373 **Outlier enrichment of McVicker's B (background selection)**

374 We used McVicker's B statistic as a measure of background selection [31]. McVicker's B
375 values that are close to 0 indicate nearly complete removal of diversity due to background
376 selection and values close to 1 indicate minimal impact. Using BEDTools [59], we
377 extracted McVicker's B values for GWAS SNPs and their matched controls. Previous
378 research suggests that McVicker's B ≤ 0.317 is a threshold for the lowest 5% of
379 McVicker's B values across the human genome [60]. Using the probability density
380 function for McVicker's B, we calculated the weighted proportion of trait-associated
381 variants with McVicker's B ≤ 0.317. Here, we used AUC statistics to quantify the
382 proportion of a SNP set with outlier values of McVicker's B. Our test of enrichment
383 involved generating a percentile rank for the trait by comparing the McVicker's B outlier
384 AUC for the trait-associated SNP set to 1000 AUC values for matched control sets of
385 SNPs (Fig. S4). High percentile ranks signify that trait-associated SNPs are enriched for
386 extreme McVicker's B statistics. Previous research has demonstrated that McVicker's B
387 statistic reliably preserves the correct rank order of SNPs [60, 61]. Nevertheless, to
388 ensure the robustness of our findings, we verified our results are consistent across more
389 stringent thresholds (McVicker's B ≤ 0.2 and 0.1), focusing European ancestry GWAS
390 hits (Table S5 and Fig. S2). We also tested whether our background selection results
391 were robust to ignoring effective size differences, once again focusing on European
392 ancestry GWAS hits (Table S6 and Fig. S3). This involved modifying the McVicker's B
393 KDE by using the same effect size for every SNP, i.e. by weighting each SNP equally.

Declarations

Ethics approval: All the GWAS summary statistic data analyzed in this paper was publicly available.

Consent for publication: Not applicable

Availability of data and materials: The GWAS summary statistics used in this paper are publicly available. Details about specific studies and SNPs used for analyses can be found in S1 Table and they are available via GitHub at https://github.com/LachanceLab/polygenic_selection. Custom code that was used to quantify polygenic signatures of natural selection is also available on GitHub at https://github.com/LachanceLab/polygenic_selection.

Competing Interests: The authors declare no competing interests.

Funding: This work was supported by an NIH MIRA grant (R35GM133727). The funder did not have any role in this article's design, analysis, or writing.

Author contributions: UH and JL conceived this study and developed the study methodology. UH curated GWAS datasets, conducted polygenic tests of selection, and

performed data visualization. JL supervised this research and provided funding. UH and JL co-wrote and edited this manuscript.

Acknowledgements: We thank Rohini Janivara, Aaron Pfennig, Mimi Holness, Ianne Lauder, King Jordan, Greg Gibson and members of the Center for Integrative Genomics at Georgia Institute of Technology for their insight and helpful comments.

References

1. Quintana-Murci L: Understanding rare and common diseases in the context of human evolution. *Genome Biol* 2016, 17:225.

2. Crespi BJ: The origins and evolution of genetic disease risk in modern humans. *Ann N Y Acad Sci* 2010, 1206:80-109.

3. Mummert A, Esche E, Robinson J, Armelagos GJ: Stature and robusticity during the agricultural transition: evidence from the bioarchaeological record. *Econ Hum Biol* 2011, 9:284-301.

4. Armstrong GL, Conn LA, Pinner RW: Trends in infectious disease mortality in the United States during the 20th century. *JAMA* 1999, 281:61-66.

5. Corbett S, Courtiol A, Lummaa V, Moorad J, Stearns S: The transition to modernity and chronic disease: mismatch and natural selection. *Nat Rev Genet* 2018, 19:419-430.

6. Bigna JJ, Noubiap JJ: The rising burden of non-communicable diseases in sub-Saharan Africa. *Lancet Glob Health* 2019, 7:e1295-e1296.

438 7. Allen L, Cobiac L, Townsend N: Quantifying the global distribution of premature
439 mortality from non-communicable diseases. *J Public Health (Oxf)* 2017, 39:698-
440 703.

441 8. Warnecke RB, Oh A, Breen N, Gehlert S, Paskett E, Tucker KL, Lurie N,
442 Rebbeck T, Goodwin J, Flack J, et al: Approaching health disparities from a
443 population perspective: the National Institutes of Health Centers for Population
444 Health and Health Disparities. *Am J Public Health* 2008, 98:1608-1615.

445 9. Visscher PM, Brown MA, McCarthy MI, Yang J: Five years of GWAS discovery.
446 *Am J Hum Genet* 2012, 90:7-24.

447 10. Lohmueller KE, Albrechtsen A, Li Y, Kim SY, Korneliussen T, Vinckenbosch N,
448 Tian G, Huerta-Sanchez E, Feder AF, Grarup N, et al: Natural selection affects
449 multiple aspects of genetic variation at putatively neutral sites across the human
450 genome. *PLoS Genet* 2011, 7:e1002326.

451 11. Tishkoff SA, Verrelli BC: Patterns of human genetic diversity: implications for
452 human evolutionary history and disease. *Annu Rev Genomics Hum Genet* 2003,
453 4:293-340.

454 12. Chheda H, Palta P, Pirinen M, McCarthy S, Walter K, Koskinen S, Salomaa V,
455 Daly M, Durbin R, Palotie A, et al: Whole-genome view of the consequences of a
456 population bottleneck using 2926 genome sequences from Finland and United
457 Kingdom. *Eur J Hum Genet* 2017, 25:477-484.

458 13. Elguero E, Delicat-Loembet LM, Rougeron V, Arnathau C, Roche B, Becquart P,
459 Gonzalez JP, Nkoghe D, Sica L, Leroy EM, et al: Malaria continues to select for
460 sickle cell trait in Central Africa. *Proc Natl Acad Sci U S A* 2015, 112:7051-7054.

461 14. Rozen R, De Braekeleer M, Daigneault J, Ferreira-Rajabi L, Gerdes M,
462 Lamoureux L, Aubin G, Simard F, Fujiwara TM, Morgan K: Cystic fibrosis
463 mutations in French Canadians: three CFTR mutations are relatively frequent in

464 a Quebec population with an elevated incidence of cystic fibrosis. *Am J Med*
465 *Genet* 1992, 42:360-364.

466 15. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, Gunes O, Hall
467 P, Hayhurst J, et al: The NHGRI-EBI GWAS Catalog: knowledgebase and
468 deposition resource. *Nucleic Acids Res* 2023, 51:D977-D985.

469 16. Torkamani A, Wineinger NE, Topol EJ: The personal and clinical utility of
470 polygenic risk scores. *Nat Rev Genet* 2018, 19:581-590.

471 17. Boyle EA, Li YI, Pritchard JK: An Expanded View of Complex Traits: From
472 Polygenic to Omnigenic. *Cell* 2017, 169:1177-1186.

473 18. Berg JJ, Coop G: A population genetic signal of polygenic adaptation. *PLoS*
474 *Genet* 2014, 10:e1004412.

475 19. Racimo F, Berg JJ, Pickrell JK: Detecting Polygenic Adaptation in Admixture
476 Graphs. *Genetics* 2018, 208:1565-1584.

477 20. Quiver MH, Lachance J: Adaptive eQTLs reveal the evolutionary impacts of
478 pleiotropy and tissue-specificity while contributing to health and disease. *HGG*
479 *Adv* 2022, 3:100083.

480 21. Mathieson I, Terhorst J: Direct detection of natural selection in Bronze Age
481 Britain. *Genome Res* 2022, 32:2057-2067.

482 22. Akbari A, Barton AR, Gazal S, Li Z, Kariminejad M, Perry A, Zeng Y, Mittnik A,
483 Patterson N, Mah M, et al: Pervasive findings of directional selection realize the
484 promise of ancient DNA to elucidate human adaptation. *bioRxiv* 2024.

485 23. Berens AJ, Cooper TL, Lachance J: The Genomic Health of Ancient Hominins.
486 *Hum Biol* 2017, 89:7-19.

24. Zeng J, de Vlaming R, Wu Y, Robinson MR, Lloyd-Jones LR, Yengo L, Yap CX, Xue A, Sidorenko J, McRae AF, et al: Signatures of negative selection in the genetic architecture of human complex traits. *Nat Genet* 2018, 50:746-753.

25. Zeng J, Xue A, Jiang L, Lloyd-Jones LR, Wu Y, Wang H, Zheng Z, Yengo L, Kemper KE, Goddard ME, et al: Widespread signatures of natural selection across human complex traits and functional genomic categories. *Nat Commun* 2021, 12:1164.

26. Koch E, Connally N, Baya N, Reeve M, Daly M, Neale B, Lander E, Bloemendal A, Sunyaev S: Genetic association data are broadly consistent with stabilizing selection shaping human common diseases and traits. *bioRxiv* 2024:2024.2006.2019.599789.

27. Patel RA, Weiss CL, Zhu H, Mostafavi H, Simons YB, Spence JP, Pritchard JK: Characterizing selection on complex traits through conditional frequency spectra. *Genetics* 2025, 229.

28. Abraham A, LaBella AL, Capra JA, Rokas A: Mosaic patterns of selection in genomic regions associated with diverse human traits. *PLoS Genet* 2022, 18:e1010494.

29. World Health Organization: Global Health Estimates 2020: deaths by cause A, sex, by country and by region, 2000–2019. WHO; 2020. 2020.

30. Johnson KE, Voight BF: Patterns of shared signatures of recent positive selection across human populations. *Nat Ecol Evol* 2018, 2:713-720.

31. McVicker G, Gordon D, Davis C, Green P: Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet* 2009, 5:e1000471.

32. 1000 Genomes Project Consortium: A global reference for human genetic variation. *Nature* 2015, 526:68.

512 33. Greaves M: Evolutionary determinants of cancer. *Cancer Discov* 2015, 5:806-
513 820.

514 34. Kim MS, Patel KP, Teng AK, Berens AJ, Lachance J: Genetic disease risks can
515 be misestimated across global populations. *Genome Biol* 2018, 19:179.

516 35. Lachance J, Tishkoff SA: SNP ascertainment bias in population genetic
517 analyses: why it is important, and how to correct it. *Bioessays* 2013, 35:780-786.

518 36. Hong EP, Park JW: Sample size and statistical power calculation in genetic
519 association studies. *Genomics Inform* 2012, 10:117-122.

520 37. Charlesworth B: Stabilizing selection, purifying selection, and mutational bias in
521 finite populations. *Genetics* 2013, 194:955-971.

522 38. Janivara R, Hazra U, Pfennig A, Harlemon M, Kim MS, Eaaswarkhanth M, Chen
523 WC, Ogunbiyi A, Kachambwa P, Petersen LN, et al: Uncovering the genetic
524 architecture and evolutionary roots of androgenetic alopecia in African men.
525 *HGG Adv* 2025, 6:100428.

526 39. Lynch M: *The origins of genome architecture.* Sunderland, Mass.: Sinauer
527 Associates; 2007.

528 40. Ohta T: The Nearly Neutral Theory of Molecular Evolution. *Annual Review of*
529 *Ecology and Systematics* 1992, 23:263-286.

530 41. Brown LM, Elbon MC, Bharadwaj A, Damle G, Lachance J: Does Effective
531 Population Size Govern Evolutionary Differences in Telomere Length? *Genome*
532 *Biol Evol* 2024, 16.

533 42. Byars SG, Voskarides K: Antagonistic Pleiotropy in Human Disease. *J Mol Evol*
534 2020, 88:12-25.

535 43. Fiziev PP, McRae J, Ulirsch JC, Dron JS, Hamp T, Yang Y, Wainschtein P, Ni Z,
536 Schraiber JG, Gao H, et al: Rare penetrant mutations confer severe risk of
537 common diseases. *Science* 2023, 380:eabo1131.

538 44. Lu T, Zhou S, Wu H, Forgetta V, Greenwood CM, Richards JB: Individuals with
539 common diseases but with a low polygenic risk score could be prioritized for rare
540 variant screening. *Genetics in Medicine* 2021, 23:508-515.

541 45. Rosenberg NA, Edge MD, Pritchard JK, Feldman MW: Interpreting polygenic
542 scores, polygenic adaptation, and human phenotypic differences. *Evol Med*
543 *Public Health* 2019, 2019:26-34.

544 46. Nagpal S, Gibson G: Dual exposure-by-polygenic score interactions highlight
545 disparities across social groups in the proportion needed to benefit. *medRxiv*
546 2024.

547 47. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB: Genetic
548 architecture: the shape of the genetic contribution to human traits and disease.
549 *Nat Rev Genet* 2018, 19:110-124.

550 48. Janivara R, Chen WC, Hazra U, Baichoo S, Agalliu I, Kachambwa P, Simonti
551 CN, Brown LM, Tambe SP, Kim MS, et al: Heterogeneous genetic architectures
552 of prostate cancer susceptibility in sub-Saharan Africa. *Nat Genet* 2024, 56:2093-
553 2103.

554 49. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ: Second-
555 generation PLINK: rising to the challenge of larger and richer datasets.
556 *Gigascience* 2015, 4:7.

557 50. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H,
558 Diekhans M, Furey TS, Harte RA, Hsu F, et al: The UCSC Genome Browser
559 Database: update 2006. *Nucleic Acids Res* 2006, 34:D590-598.

560 51. Pers TH, Timshel P, Hirschhorn JN: SNPsnap: a Web-based tool for
561 identification and annotation of matched SNPs. *Bioinformatics* 2015, 31:418-420.

562 52. Carvalho NRG, Harris AM, Lachance J: Different genetic architectures of
563 complex traits and their relevance to polygenic score performance. *bioRxiv*
564 2022:2022.2010.2029.514295.

565 53. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P,
566 Lander ES, Lubitz SA, Ellinor PT, Kathiresan S: Genome-wide polygenic scores
567 for common diseases identify individuals with risk equivalent to monogenic
568 mutations. *Nat Genet* 2018, 50:1219-1224.

569 54. Graham BE, Plotkin B, Muglia L, Moore JH, Williams SM: Estimating prevalence
570 of human traits among populations from polygenic risk scores. *Human genomics*
571 2021, 15:1-16.

572 55. Lipson M, Loh PR, Patterson N, Moorjani P, Ko YC, Stoneking M, Berger B,
573 Reich D: Reconstructing Austronesian population history in Island Southeast
574 Asia. *Nat Commun* 2014, 5:4689.

575 56. Pickrell JK, Pritchard JK: Inference of population splits and mixtures from
576 genome-wide allele frequency data. *PLoS Genet* 2012, 8:e1002967.

577 57. Pedregosa F, Varoquaux, G. , Gramfort, A. , Michel, V. , Thirion, B. , Grisel, O. ,
578 Blondel, M. , Prettenhofer, P. , Weiss, R. , Dubourg, V. , Vanderplas, J. , Passos,
579 A. , Cournapeau, D. a, Brucher, M. , Perrot, M. ,Duchesnay, E.: Scikit-learn:
580 Machine Learning in Python. *Journal of Machine Learning Research* 2011,
581 12:2825--2830.

582 58. Voight BF, Kudaravalli S, Wen X, Pritchard JK: A map of recent positive selection
583 in the human genome. *PLoS Biol* 2006, 4:e72.

584 59. Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic
585 features. *Bioinformatics* 2010, 26:841-842.

586 60. Torres R, Szpiech ZA, Hernandez RD: Human demographic history has amplified
587 the effects of background selection across the genome. *PLoS Genet* 2018,
588 14:e1007387.

589 61. Comeron JM: Background selection as baseline for nucleotide variation across
590 the Drosophila genome. *PLoS Genet* 2014, 10:e1004434.

591 62. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, Weeks EM, Wang
592 M, Hindy G, Zhou W, et al: Discovery and systematic characterization of risk
593 variants and genes for coronary artery disease in over a million participants. *Nat*
594 *Genet* 2022, 54:1803-1815.

595 63. Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kawakami E, Sugishita H, Sakaue
596 S, Matoba N, Low SK, Okada Y, et al: Large-scale genome-wide association
597 study in a Japanese population identifies novel susceptibility loci across different
598 diseases. *Nat Genet* 2020, 52:669-679.

599 64. Tcheandjieu C, Zhu X, Hilliard AT, Clarke SL, Napolioni V, Ma S, Lee KM, Fang
600 H, Chen F, Lu Y, et al: Large-scale genome-wide association study of coronary
601 artery disease in genetically diverse populations. *Nat Med* 2022, 28:1679-1692.

602 65. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-
603 Jacobs L, Giese AK, van der Laan SW, Gretarsdottir S, et al: Multiancestry
604 genome-wide association study of 520,000 subjects identifies 32 loci associated
605 with stroke and stroke subtypes. *Nat Genet* 2018, 50:524-537.

606 66. Mishra A, Malik R, Hachiya T, Jurgenson T, Namba S, Posner DC, Kamanu FK,
607 Koido M, Le Grand Q, Shi M, et al: Stroke genetics informs drug discovery and
608 risk prediction across ancestries. *Nature* 2022, 611:115-123.

609 67. Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL,
610 Jackson VE, Shrine N, Qiao D, Bartz TM, Kim DK, et al: Genetic landscape of
611 chronic obstructive pulmonary disease identifies heterogeneous cell-type and
612 phenotype associations. *Nat Genet* 2019, 51:494-505.

613 68. Shrine N, Izquierdo AG, Chen J, Packer R, Hall RJ, Guyatt AL, Batini C,
614 Thompson RJ, Pavuluri C, Malik V, et al: Multi-ancestry genome-wide
615 association analyses improve resolution of genes and pathways influencing lung
616 function and chronic obstructive pulmonary disease risk. *Nat Genet* 2023,
617 55:410-422.

618 69. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC,
619 Caporaso NE, Johansson M, Xiao X, Li Y, et al: Large-scale association analysis
620 identifies new lung cancer susceptibility loci and heterogeneity in genetic
621 susceptibility across histological subtypes. *Nat Genet* 2017, 49:1126-1132.

622 70. Byun J, Han Y, Li Y, Xia J, Long E, Choi J, Xiao X, Zhu M, Zhou W, Sun R, et al:
623 Cross-ancestry genome-wide meta-analysis of 61,047 cases and 947,237
624 controls identifies new susceptibility loci contributing to lung cancer. *Nat Genet*
625 2022, 54:1167-1177.

626 71. Bellenguez C, Kucukali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, Naj
627 AC, Campos-Martin R, Grenier-Boley B, Andrade V, et al: New insights into the
628 genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* 2022,
629 54:412-436.

630 72. Cai L, Wheeler E, Kerrison ND, Luan J, Deloukas P, Franks PW, Amiano P,
631 Ardanaz E, Bonet C, Fagherazzi G, et al: Genome-wide association analysis of
632 type 2 diabetes in the EPIC-InterAct study. *Sci Data* 2020, 7:393.

633 73. Spracklen CN, Horikoshi M, Kim YJ, Lin K, Bragg F, Moon S, Suzuki K, Tam
634 CHT, Tabara Y, Kwak SH, et al: Identification of type 2 diabetes loci in 433,540
635 East Asian individuals. *Nature* 2020, 582:240-245.

636 74. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, Yu GZ,
637 Rueger S, Speidel L, Kim YJ, et al: Multi-ancestry genetic study of type 2
638 diabetes highlights the power of diverse populations for discovery and
639 translation. *Nat Genet* 2022, 54:560-572.

640 75. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, Tin A, Wang L, Chu AY,
641 Hoppmann A, et al: A catalog of genetic loci associated with kidney function from
642 analyses of a million individuals. *Nat Genet* 2019, 51:957-972.

643 76. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N,
644 Ikegawa S, Hirata M, Matsuda K, et al: Genetic analysis of quantitative traits in
645 the Japanese population links cell types to complex human diseases. *Nat Genet*
646 2018, 50:390-400.

647 77. Surendran P, Feofanova EV, Lahrouchi N, Ntalla I, Karthikeyan S, Cook J, Chen
648 L, Mifsud B, Yao C, Kraja AT, et al: Discovery of rare variants associated with
649 blood pressure regulation through meta-analysis of 1.3 million individuals. *Nat*
650 *Genet* 2020, 52:1314-1332.

651 78. Giri A, Hellwege JN, Keaton JM, Park J, Qiu C, Warren HR, Torstenson ES,
652 Kovesdy CP, Sun YV, Wilson OD, et al: Trans-ethnic association study of blood
653 pressure determinants in over 750,000 individuals. *Nat Genet* 2019, 51:51-62.

654 79. Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-
655 Tajes J, Farrington S, Svinti V, Palles C, Orlando G, et al: Association analyses
656 identify 31 new risk loci for colorectal cancer susceptibility. *Nat Commun* 2019,
657 10:2154.

80. Lu Y, Kweon SS, Cai Q, Tanikawa C, Shu XO, Jia WH, Xiang YB, Huyghe JR, Harrison TA, Kim J, et al: Identification of Novel Loci and New Risk Variant in Known Loci for Colorectal Cancer Risk in East Asians. *Cancer Epidemiol Biomarkers Prev* 2020, 29:477-486.

81. Fernandez-Rozadilla C, Timofeeva M, Chen Z, Law P, Thomas M, Schmit S, Diez-Obrero V, Hsu L, Fernandez-Tajes J, Palles C, et al: Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and east Asian ancestries. *Nat Genet* 2023, 55:89-99.

82. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen TH, Wang Q, Bolla MK, et al: Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 2019, 104:21-34.

83. Shu X, Long J, Cai Q, Kweon SS, Choi JY, Kubo M, Park SK, Bolla MK, Dennis J, Wang Q, et al: Identification of novel breast cancer susceptibility loci in meta-analyses conducted among Asian and European descendants. *Nat Commun* 2020, 11:1217.

674 # Tables

675

Disease	Ancestry of GWAS SNP set [citation]	No. of SNPs analyzed	PolyGraph q-value	Max iHS %ile (population)	McVicker's B statistic %ile
Ischemic Heart Disease	European [62]	200	0.1689	98 (ITU)	98.8
	East Asian [63]	829	5.21×10^{-11}	>99.9 (TSI)	>99.9
	Multi-ancestry [64]	50	0.6884	86 (PEL)	90.3
Stroke	European [65]	263	0.2951	88 (YRI)	98.8
	East Asian [63]	122	0.0941	84 (GWD)	95.6
	Multi-ancestry [66]	123	0.6884	74 (IBS)	97.1
COPD	European [67]	79	0.9245	93 (PUR)	98.8
	East Asian [63]	92	0.7000	95 (JPT)	>99.9
	Multi-ancestry [68]	100	0.6884	89 (GIH, PUR)	99.8
Lung Cancer	European [69]	258	0.9245	92 (GBR)	91.8
	East Asian [63]	101	0.1144	92 (GBR)	74.9
	Multi-ancestry [70]	124	4.37×10^{-9}	>99.99 (PEL)	39.8
Alzheimer's Disease	European [71]	75	0.8876	94.5 (TSI)	96.1
Type 2 Diabetes	European [72]	183	0.4697	85 (JPT)	97.7
	East Asian [73]	158	0.1777	98 (CLM)	94.1
	Multi-ancestry [74]	1661	0.6884	99 (PEL)	99.9
Chronic Kidney Disease	European [75]	2283	0.0113	98 (CDX, PEL)	>99.9
	East Asian [76]	583	0.0010	57 (CHB)	>99.9
	Multi-ancestry [75]	521	0.3186	70 (FIN, GWD)	>99.9
Hypertensive Heart Disorder	European [77]	1124	0.0160	99 (PEL)	>99.9
	East Asian [76]	315	0.0100	96 (ESN)	>99.9
	Multi-ancestry [78]	226	0.6884	90 (CHS)	99.9
Colon Cancer	European [79]	83	0.9245	78 (GIH)	88.0
	East Asian [80]	60	0.0061	72 (MSL)	99.9
	Multi-ancestry [81]	188	0.7225	62 (KHV)	99.7
Breast Cancer	European [82]	242	0.9245	89 (CHS)	99.2
	East Asian [63]	118	0.0265	81 (CDX)	78.6
	Multi-ancestry [83]	100	0.9716	94 (CHS)	98.9

676

677 **Table 1. Overview of the diseases, SNP sets, and polygenic tests of selection in**
678 **this paper.** Genetic ancestries and citations for each disease-associated SNP set are
679 listed here, as are the number independent ($r^2 < 0.2$) disease-associated SNPs. FDR-
680 adjusted q-values from PolyGraph analyses are list here (as per Fig. 2). Percentile ranks

681 refer to outlier enrichment comparisons between trait-associated SNP sets and 1000
682 matched control SNP sets. For each combination of disease and GWAS ascertainment
683 scheme, the maximum integrated haplotype score (iHS) percentile rank is listed (as per
684 Fig. 3), as is the relevant 1000 Genomes Project population code. Percentile ranks of
685 McVicker's B statistics quantify enrichment of signatures of background selection (as per
686 Fig. 4). Alzheimer's data were not available for East Asian and multi-ancestry GWAS.
687 Additional details about each GWAS SNP set are described in Table S1.

Figures

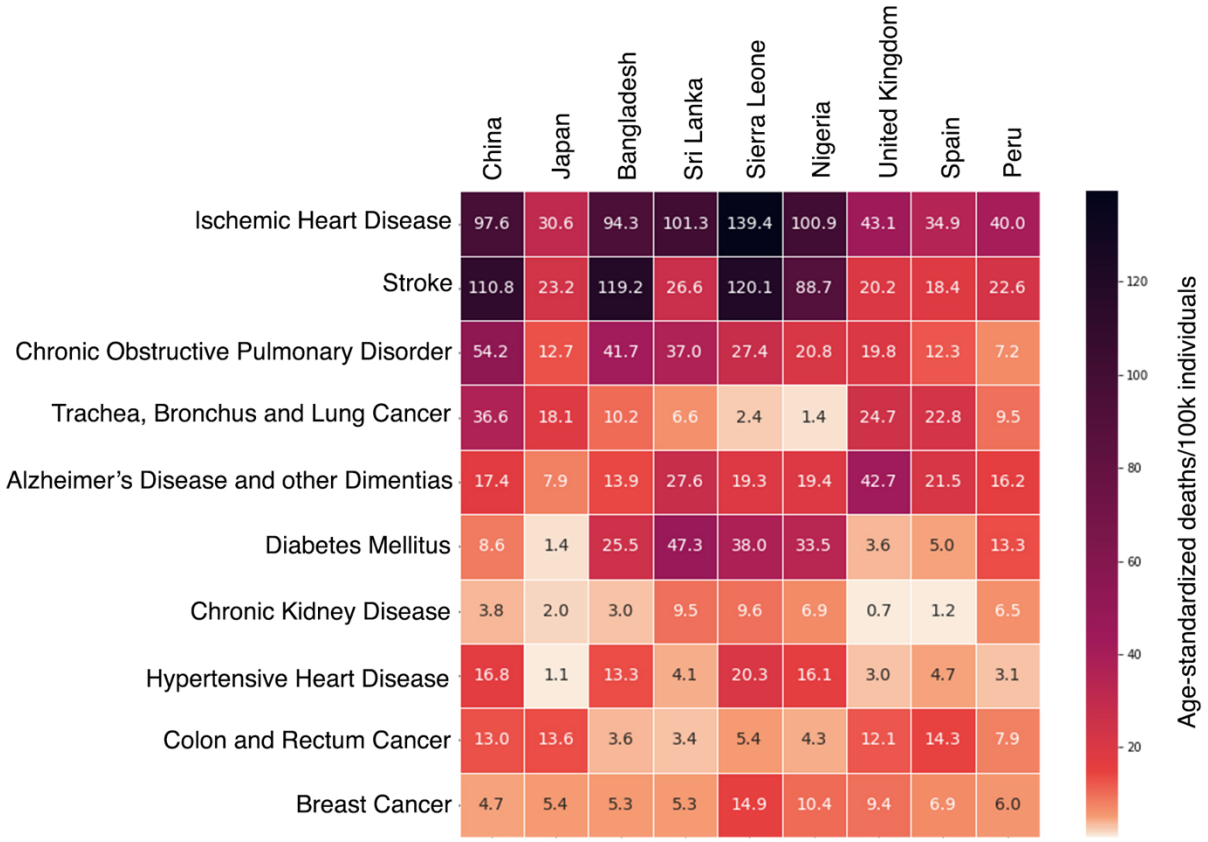

Fig. 1. Heatmap of age-standardized mortality rates per 100,000 individuals for the 10 hereditary diseases with the largest global burden. Source of age-adjusted mortality rates: 2020 World Health Organization (WHO) Report [29]. Countries shown here have matching populations in the 1000 Genomes Project (1KGP).

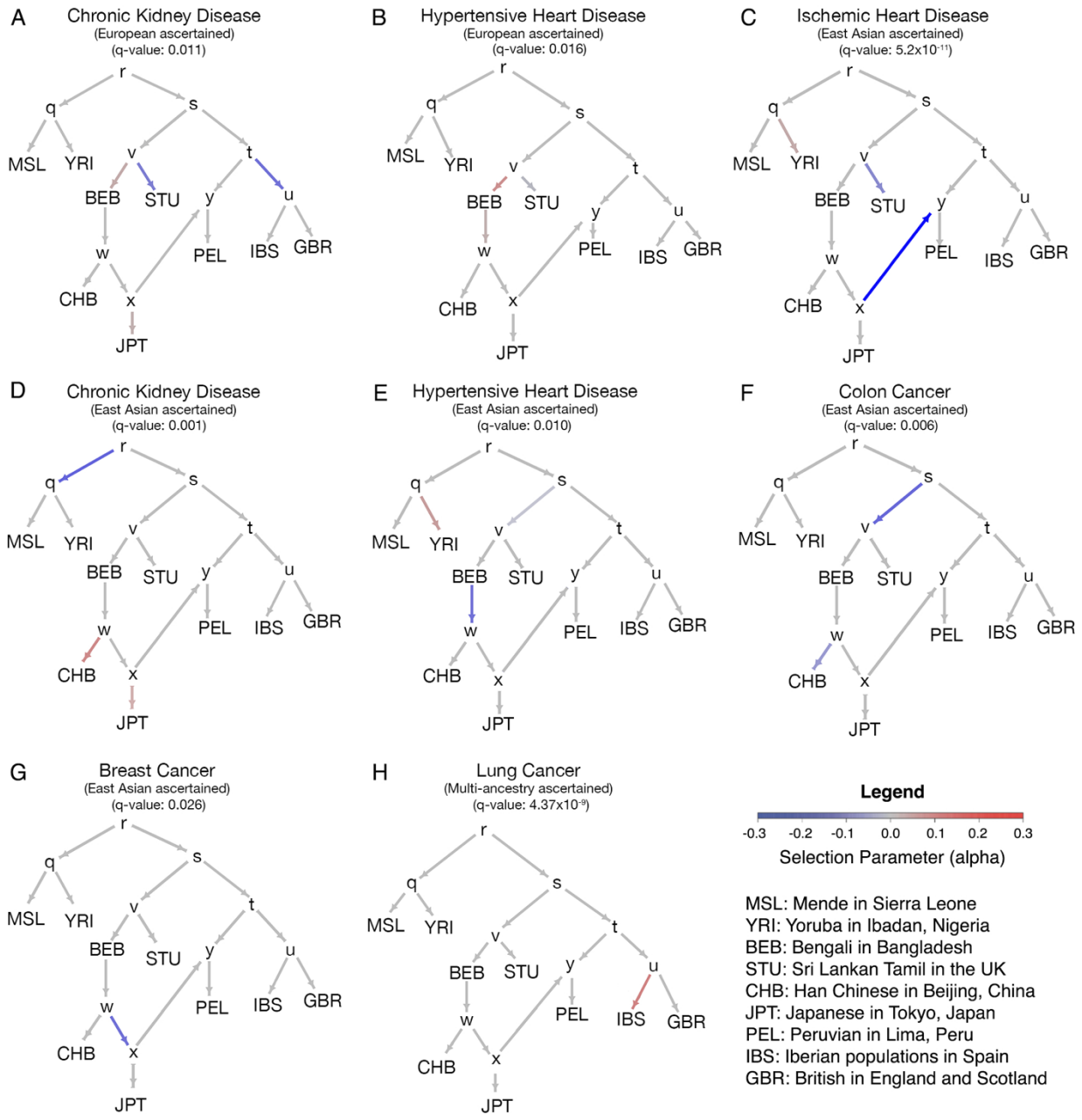

Fig. 2. PolyGraph results do not show consistent evidence of polygenic adaptation across multiple GWAS ascertainment schemes. Note that 20 out of the 28 disease-GWAS ancestry combinations described in Table 1 did not show any evidence of polygenic selection – only the exceptions are plotted here. Allele frequencies from the 1KGP were used here, and population codes are indicated in the legend. q-values are FDR-adjusted. The selection parameter alpha reports the product of the selection coefficient for advantageous alleles and the duration of the selective process for each branch.

Disease	GWAS Ascertainment	African							South Asian					East Asian					European					Admixed American			
		ACB	ESN	GWD	LWK	MSL	YRI	ASW	BEB	GIH	ITU	PJL	STU	CDX	CHB	CHS	JPT	KHV	CEU	FIN	GBR	IBS	TSI	CLM	MXL	PEL	PUR
Ischemic Heart Disease	European	28	48	35	55	84	67	90	97	89	98	85	96	31	83	77	18	75	8	6	60	42	69	33	79	87	84
	East Asian	23	20	59	9	53	5	21	77	46	70	44	68	77	91	81	61	53	62	35	38	88	> 99	71	55	27	54
	Multi-ancestry	29	74	50	40	44	62	28	48	30	26	40	52	39	35	23	57	23	5	34	35	40	17	31	46	86	15
Stroke	European	79	71	76	73	70	88	78	39	13	40	28	53	65	79	42	74	35	48	49	45	71	57	14	43	11	52
	East Asian	46	50	84	15	23	45	68	78	77	69	30	58	82	68	57	55	60	73	14	66	54	59	13	38	73	27
	Multi-ancestry	11	15	28	39	30	32	20	1	42	11	42	43	39	16	45	18	33	11	66	66	74	56	23	58	57	49
Chronic Obstructive Pulmonary Disorder	European	5	23	31	53	20	48	36	26	65	10	42	59	4	4	2	2	13	75	38	75	88	16	47	34	10	93
	East Asian	40	23	16	58	37	40	72	59	60	53	19	51	59	10	89	95	26	82	69	82	85	77	79	85	94	45
	Multi-ancestry	17	40	50	6	11	44	51	86	89	84	65	86	79	35	67	29	56	77	62	78	70	82	60	36	75	89
Lung Cancer	European	67	41	63	38	65	69	36	72	69	43	58	48	54	50	37	80	43	90	78	92	36	78	88	70	81	63
	East Asian	57	72	72	48	26	66	5	26	41	58	64	55	60	39	39	75	41	52	20	92	45	84	46	68	78	32
	Multi-ancestry	98	95	84	30	90	96	74	50	38	29	14	26	26	23	14	67	28	92	42	73	75	91	76	43	> 99	47
Alzheimer's Disease	European	37	41	41	40	38	46	1	51	14	82	49	90	54	84	39	34	91	57	43	64	88	94	10	77	49	62
	East Asian	No Data																									
	Multi-ancestry	No Data																									
Type 2 Diabetes	European	18	49	10	70	21	41	18	56	49	64	62	59	8	26	84	85	38	23	77	64	29	38	75	65	35	63
	East Asian	15	14	62	52	25	56	70	82	92	95	91	59	85	61	84	88	91	60	49	78	28	56	98	89	94	78
	Multi-ancestry	6	5	14	8	4	11	2	44	26	43	68	40	44	14	59	23	50	26	24	44	10	38	59	18	99	24
Chronic Kidney Disease	European	8	37	34	30	23	67	20	33	12	8	66	19	98	18	79	11	25	11	82	42	7	6	42	74	98	25
	East Asian	8	15	18	2	7	12	0	11	15	21	5	11	17	57	52	40	14	32	56	20	37	31	46	17	30	34
	Multi-ancestry	13	18	70	34	35	29	9	1	6	1	1	15	2	2	1	2	1	28	70	44	6	14	16	59	49	8
Hypertensive Heart Disease	European	93	82	91	41	85	93	74	42	16	16	68	87	23	14	28	12	77	36	33	56	86	74	53	83	99	46
	East Asian	89	96	55	53	88	91	43	75	67	61	83	31	87	82	68	77	94	61	48	82	78	76	32	28	58	64
	Multi-ancestry	81	60	40	71	80	76	68	36	51	59	65	30	70	70	90	51	88	26	52	28	50	60	57	46	79	65
Colon Cancer	European	5	33	25	29	6	7	13	13	78	50	50	3	41	37	22	37	35	45	51	41	60	34	24	15	25	26
	East Asian	51	16	44	7	72	45	12	7	50	22	12	12	13	9	6	14	2	56	60	16	18	62	9	13	2	1
	Multi-ancestry	6	3	9	31	15	2	26	19	45	12	26	2	53	38	21	58	62	8	19	10	36	57	27	16	15	53
Breast Cancer	European	27	39	13	46	2	6	25	6	6	1	4	0	56	54	89	51	44	27	57	1	8	10	29	20	76	18
	East Asian	60	71	47	35	26	56	34	59	33	58	38	28	81	20	74	31	74	35	65	25	39	42	43	50	44	31
	Multi-ancestry	65	37	26	7	15	9	18	63	58	68	34	78	90	85	94	93	80	17	58	45	5	3	46	62	74	70

0 20 40 60 80 100

Trait percentile rank compared to control SNPs

Fig. 3. Sparse signals of recent positive selection acting on complex diseases. Plotted here are results from SNPs sets that were ascertained in European, East Asian, and multi-ancestry GWAS. Percentile ranks quantify how much disease-associated loci are enriched for outlier values of integrative haplotype scores (|iHS| >1.96) compared to 1000 sets of control SNPs, with darker shading indicating stronger polygenic signatures of recent positive selection. Population codes, i.e., column headings, are from the 1000 Genomes Project (1KGP).

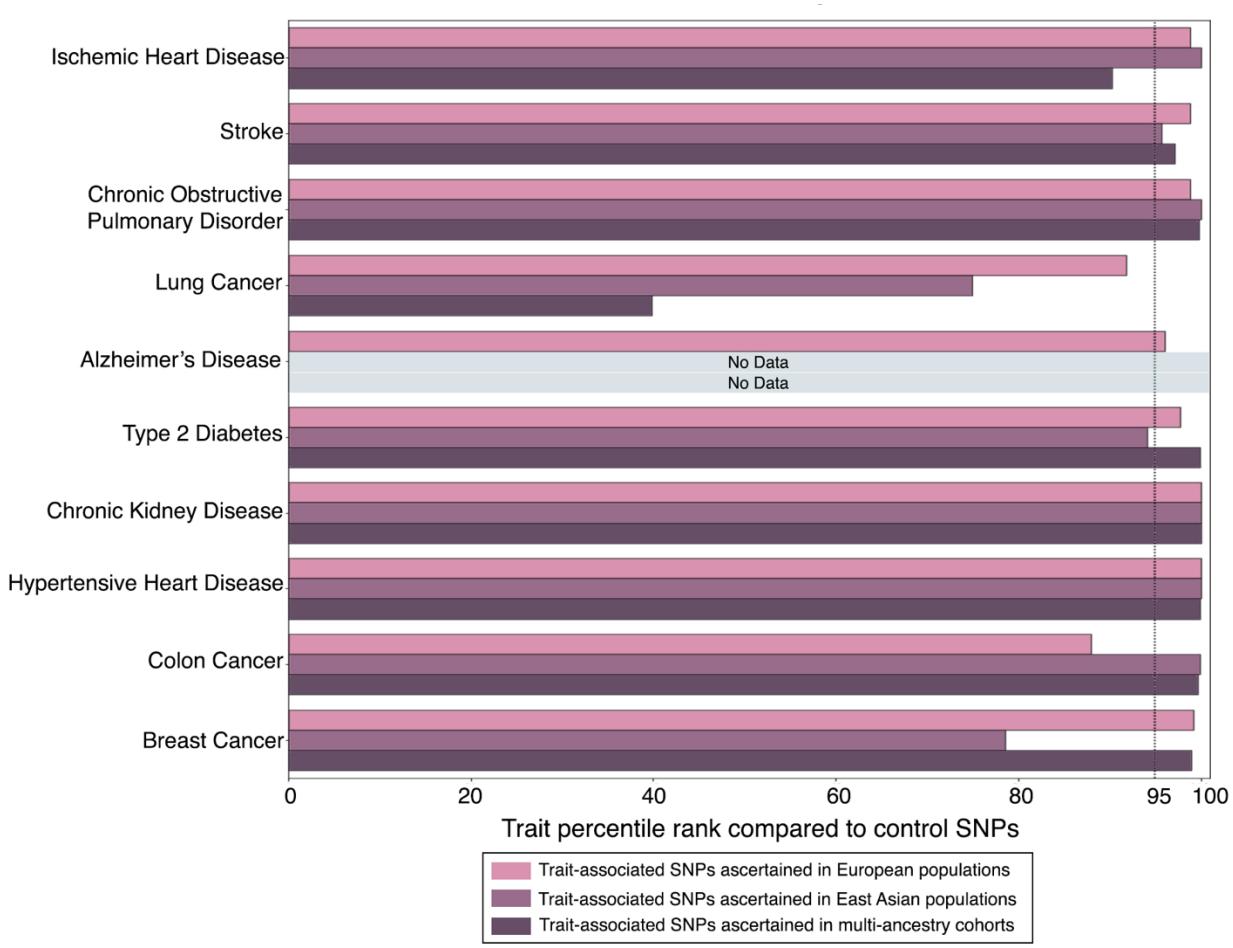

Fig. 4. Disease-associated SNP sets are enriched for signatures background selection. Results from SNP sets that were ascertained in European, East Asian, and multi-ancestry GWAS are plotted here (exact percentile ranks can be found in Table 1). Percentile ranks for each disease involve comparison between sets of disease-associated to 1000 sets of control SNPs. The dotted line marks the 95th percentile of control SNP sets. As per [60], a McVicker's B statistic outlier threshold of 0.317 was used.

721 ## Additional files

722

723 **Additional file 1.** Tables S1-S6 are embedded in a tabbed .xslx file. **Table S1**: Additional
724 details about GWAS SNP sets. Details include citations, the genetic ancestry of the each
725 GWAS study cohort, sample sizes, number of SNPs before and after LD clumping, and
726 links to SNP lists (including iHS statistics for each SNP). **Table S2**: Branch statistics for
727 PolyGraph analyses. Summary statistics are shown for ten complex diseases and three
728 different GWAS ascertainment schemes: European, East Asian, and multi-ancestry
729 cohorts. Branch-specific and admixture graph-wide statistics are included. **Table S3**:
730 Exact iHS outlier enrichment percentile rank statistics for multiple ascertainment schemes
731 and 26 global populations from the 1KGP. Higher percentile ranks indicate enrichment
732 for signatures of recent positive selection. **Table S4**: Effects of weighting the KDE by
733 effect size for iHS statistics. Outlier enrichment percentile ranks are for European
734 ascertained SNP sets.Higher percentile ranks indicate enrichment for signatures of recent
735 positive selection. **Table S5**: Comparisons of different outlier thresholds for McVicker's B
736 statistics. Outlier enrichment percentile ranks are for European ascertained SNP sets.
737 Higher percentile ranks indicate enrichment for signatures of background selection. **Table**
738 **S6**: Effects of weighting the KDE by effect size for McVicker's B statistics. Outlier
739 enrichment percentile ranks are for European ascertained SNP sets. Higher percentile
740 ranks indicate enrichment for signatures of background selection.

741

742 **Additional file 2**: Figures S1-S4 are combined in a singled merged file.