Current clinical use of polygenic scores will risk exacerbating health disparities

Alicia R. Martin[1,2,3], Masahiro Kanai[1,2,3,4,5], Yoichiro Kamatani[5,6], Yukinori Okada[5,7,8], Benjamin M. Neale[1,2,3], Mark J. Daly[1,2,3,9]

[1] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA
[2] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[3] Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[4] Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA
[5] Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan
[6] Kyoto-McGill International Collaborative School in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan
[7] Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan
[8] Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita 565-0871, Japan
[9] Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

Correspondence: Alicia R. Martin, armartin@broadinstitute.org

Abstract

Polygenic risk scores (PRS) are poised to improve biomedical outcomes via precision medicine. However, the major *ethical* and *scientific* challenge surrounding clinical implementation is that they are many-fold more accurate in European ancestry individuals than others. This disparity is an inescapable consequence of Eurocentric genome-wide association study biases. This highlights that—unlike clinical biomarkers and prescription drugs, which may individually work better in some populations but do not ubiquitously perform far better in European populations—clinical uses of PRS today would systematically afford greater improvement to European descent populations. Early diversifying efforts show promise in levelling this vast imbalance, even when non-European sample sizes are considerably smaller than the largest studies to date. To realize the full and equitable potential of PRS, we must prioritize greater diversity in genetic studies and public dissemination of summary statistics to ensure that health disparities are not increased for those already most underserved.

Keywords: health disparities, genetic risk prediction, polygenic risk scores, diversity, population genetics, statistical genetics

Polygenic risk scores (PRS), which predict complex traits using genetic data, are of burgeoning interest to the clinical community as researchers demonstrate their growing power to improve clinical care, genetic studies of a wide range of phenotypes increase in size and power, and genotyping costs plummet to less than US$50. Many earlier

50 criticisms of limited prediction power are now recognized to have been chiefly an issue
51 of insufficient sample size, which is no longer the case for many outcomes[1]. For
52 example, polygenic risk scores alone already predict breast cancer, prostate cancer,
53 and type 1 diabetes risk in European descent patients more accurately than current
54 clinical models[2-4]. Additionally, integrated models of PRS together with other lifestyle
55 and clinical factors have enabled clinicians to more accurately quantify the risk of heart
56 attack for patients; consequently, they have more effectively targeted the reduction of
57 LDL cholesterol and by extension heart attack by prescribing statins to patients at the
58 greatest overall risk of cardiovascular disease[5-9]. Promisingly, return of genetic risk of
59 complex disease to at-risk patients does not induce significant self-reported negative
60 behavior or psychological function, and some potentially positive behavioral changes
61 have been detected[10]. While we share enthusiasm about the potential of PRS to
62 improve health outcomes through their eventual routine implementation as clinical
63 biomarkers, we consider the consistent observation that they are currently of far greater
64 predictive value in individuals of recent European descent than in others to be the major
65 *ethical* and *scientific* challenge surrounding clinical translation and, at present, the most
66 critical limitation to genetics in precision medicine. The scientific basis of this imbalance
67 has been demonstrated theoretically, in simulations, and empirically across many traits
68 and diseases[11-22].
69
70 All studies to date using well-powered genome-wide association studies (GWAS) to
71 assess the predictive value of PRS across a range of traits and populations have made
72 a consistent observation: PRS predict individual risk far more accurately in Europeans

73 than non-Europeans[15,16,18-24] . Rather than chance or biology, this is a predictable
74 consequence of the fact that the genetic discovery efforts to date heavily
75 underrepresent non-European populations globally. The correlation between true and
76 genetically predicted phenotypes decays with genetic divergence from the makeup of
77 the discovery GWAS, meaning that the accuracy of polygenic scores in different
78 populations is highly dependent on the study population representation in the largest
79 existing 'training' GWAS. Here, we document study biases that underrepresent non-
80 European populations in current GWAS, and explain the fundamental concepts
81 contributing to reduced phenotypic variance explained with increasing genetic
82 divergence from populations included in GWAS.
83
84 **Predictable basis of disparities in PRS accuracy**
85 Poor generalizability of genetic studies across populations arises from the
86 overwhelming abundance of European descent studies and dearth of well-powered
87 studies in globally diverse populations[25-28]. According to the GWAS catalog, ~79% of all
88 GWAS participants are of European descent despite making up only 16% of the global
89 population (**Figure 1**). This is especially problematic as previous studies have shown
90 that Hispanic/Latino and African American studies contribute an outsized number of
91 associations relative to studies of similar sizes in Europeans[27]. More concerningly, the
92 fraction of non-European individuals in GWAS has stagnated or declined since late
93 2014 (**Figure 1**), suggesting that we are not on a trajectory to correct this imbalance.
94 These numbers provide a composite metric of study availability, accessibility, and use—
95 cohorts that have been included in numerous GWAS are represented multiple times,

which may disproportionately include cohorts of European descent. However, whereas the average sample sizes of GWAS in Europeans continue to grow, they have stagnated and remain several-fold smaller in other populations (**Supplementary Figure 1**).

The relative sample compositions of GWAS result in highly predictable disparities in prediction accuracy; population genetics theory predicts that genetic risk prediction accuracy will decay with increasing genetic divergence between the original GWAS sample and target of prediction, a function of population history[13,14]. This pattern can be attributed to several statistical observations which we detail below: 1) GWAS favor the discovery of genetic variants that are common in the study population; 2) linkage disequilibrium (LD) differentiates marginal effect size estimates for polygenic traits across populations, even when causal variants are the same; and 3) environment and demography differ across populations. Notably, the first two phenomena degrade prediction performance across populations substantially even when there exist no biological, environmental, or diagnostic differences, whereas the environment and demography may interact to drive differential forces of natural selection that in turn drive differences in causal genetic architecture. (We define the causal genetic architecture as the true effects of variants that impact a phenotype that would be identified in a population of infinite sample size. Unlike effect size estimates, true effects are typically modeled as invariant with respect to LD and allele frequency differences across populations.)

119 *Common discoveries and low-hanging fruit*

120 First, the power to discover an association in a genetic study depends on the effect size
121 and frequency of the variant[29]. This dependence means that the most significant
122 associations tend to be more common in the populations in which they are discovered
123 than elsewhere[13,30]. For example, GWAS catalog variants are more common on
124 average in European populations compared to East Asian and African populations
125 (**Figure 2B**), an observation not representative of genomic variants at large.
126 Understudied populations offer low-hanging fruit for genetic discovery because variants
127 that are common in these groups but rare or absent in European populations could not
128 be discovered even with very large European sample sizes. Some examples include
129 *SLC16A11* and *HNF1A* associations with type II diabetes in Latino populations, as well
130 as *APOL1* associations with end-stage kidney disease and associations with prostate
131 cancer in African descent populations[31-34]. If we assume that causal genetic variants
132 have an equal effect across all populations—an assumption with some empirical
133 support that offers the best case scenario for transferability[35-40]—Eurocentric GWAS
134 biases mean that variants associated with risk are disproportionately common and
135 discovered in European populations, accounting for a larger fraction of the phenotypic
136 variance there[13]. Furthermore, imputation reference panels share the same study
137 biases as in GWAS[41], creating challenges for imputing sites that are rare in European
138 populations but common elsewhere when the catalog of non-European haplotypes is
139 substantially smaller. These issues are insurmountable through statistical methods
140 alone[13], but rather motivate substantial investments in more diverse populations to
141 produce similar-sized GWAS of biomedical phenotypes in other populations.

142

143 *Linkage disequilibrium*

144 Second, LD, the correlation structure of the genome, varies across populations due to
145 demographic history (**Figure 2A,C-E**). These LD differences in turn drive differences in
146 effect size estimates (i.e. predictors) from GWAS across populations in proportion to LD
147 between tagging and causal SNP pairs, even when causal effects are the same [35,37-40]
148 (**Supplementary Note**). Differences in effect size estimates due to LD differences may
149 typically be small for most regions of the genome (**Figure 2C-E**), but PRS sum across
150 these effects, also aggregating these population differences. While it would be ideal to
151 use causal effects rather than correlated effect size estimates to calculate PRS, it may
152 not be feasible to fine-map most variants to a single locus to solve issues of low
153 generalizability, even with very large GWAS. This is because complex traits are highly
154 polygenic, meaning most of our prediction power comes from small effects that do not
155 meet genome-wide significance and/or cannot be fine-mapped, even in many of the
156 best-powered GWAS to date[42].

157

158 *Complexities of history, selection, and the environment*

159 Lastly, other cohort considerations may further worsen prediction accuracy differences
160 across populations in less predictable ways. GWAS ancestry study biases and LD
161 differences across populations are extremely challenging to address, but these issues
162 actually make many favorable assumptions that all causal loci have the same impact
163 and are under equivalent selective pressure in all populations. In contrast, other effects
164 on polygenic adaptation or risk scores such as long-standing environmental differences

165 across global populations that have resulted in differing responses of natural selection
166 can impact populations differently based on their unique histories. Additionally, residual
167 uncorrected population stratification may impact risk prediction accuracy across
168 populations, but the magnitude of its effect is currently unclear. These effects are
169 particularly challenging to disentangle, as has clearly been demonstrated for height,
170 where evidence of polygenic adaptation and/or its relative magnitude is under
171 question[43,44]. Comparisons of geographically stratified phenotypes like height across
172 populations with highly divergent genetic backgrounds and mean environmental
173 differences, such as differences in resource abundance during development across
174 continents, are especially prone to confounding from correlated environmental and
175 genetic divergence[43,44]. This residual stratification can lead to over-predicted differences
176 across geographical space[45].
177
178 Related to stratification, most PRS methods do not explicitly address recent admixture
179 and none consider recently admixed individuals' unique local mosaic of ancestry; further
180 methods development is needed. Additionally, comparing PRS across environmentally
181 stratified cohorts, such as in some biobanks with healthy volunteer effects versus
182 disease study datasets or hospital-based cohorts, requires careful consideration of
183 technical differences, collider bias, as well as variability in baseline health status among
184 studies. It is also important to consider differences in definitions of clinical phenotypes
185 and heterogeneity of sub-phenotypes among countries.
186

187 Differences in environmental exposure, gene-gene interactions, gene-environment
188 interactions, historical population size dynamics, statistical noise, some potential causal
189 effect differences, and/or other factors will further limit generalizability for genetic risk
190 scores in an unpredictable, trait-specific fashion[46-49]. Complex traits do not behave in a
191 genetically deterministic manner, with some environmental factors dwarfing individual
192 genetic effects, creating outsized issues of comparability across globally diverse
193 populations. Among psychiatric disorders for example, whereas schizophrenia has a
194 nearly identical genetic basis across East Asians and Europeans (r_g=0.98)[40],
195 substantially different rates of alcohol use disorder across populations are partially
196 explained by differences in availability and genetic differences impacting alcohol
197 metabolism[50]. While non-linear genetic factors explain little variation in complex traits
198 beyond a purely additive model[51], some unrecognized nonlinearities and gene-gene
199 interactions can also induce genetic risk prediction challenges, as pairwise interactions
200 are likely to vary more across populations than individual SNPs. Mathematically, we can
201 simplistically think of this in terms of a two-SNP model, in which the sum of two SNP
202 effects is likely to explain more phenotypic variance than the product of the same SNPs.
203 Some machine learning approaches may thus modestly improve PRS accuracy beyond
204 current approaches for some phenotypes[52], but most likely for atypical traits with simpler
205 architectures, known interactions, and poor prediction generalizability across
206 populations, such as skin pigmentation[53].
207
208 **Limited generalizability of PRS across diverse populations**

209 So far, multi-ethnic work has been slow in most disease areas[54], limiting even the
210 opportunity to assess PRS in non-European cohorts. Nonetheless, some previous work
211 has assessed prediction accuracy across diverse populations in several traits and
212 diseases for which GWAS summary statistics are available and identified large
213 disparities across populations (**Supplementary Note**). These disparities are not simply
214 methodological issues, as various approaches (e.g. pruning and thresholding versus
215 LDPred) and accuracy metrics (R^2 for quantitative traits and various pseudo-R^2 metrics
216 for binary traits) illustrate this consistently poorer performance in populations distinct
217 from discovery samples across a range of polygenic traits (**Supplementary Table 1**).
218 These assessments are becoming increasingly feasible with the growth and public
219 availability of global biobanks as well as diversifying priorities from funding
220 agencies[55,56]. We assessed how prediction accuracy decayed across globally diverse
221 populations for 17 anthropometric and blood panel traits in the UK Biobank (UKBB)
222 when using European-derived summary statistics (**Supplementary Note**). Consistent
223 with previous studies, we find that relative to European prediction accuracy, genetic
224 prediction accuracy was far lower in other populations: 1.6-fold lower in Hispanic/Latino
225 Americans, 1.7-fold lower in South Asians, 2.5-fold lower in East Asians, and 4.9-fold
226 lower in Africans on average (**Figure 3**).
227

228 **Prioritizing diversity shows early promise for PRS**

229 Early diversifying GWAS efforts have been especially productive for informing on
230 questions surrounding risk prediction. Rather than varying the prediction target dataset,
231 some GWAS in diverse populations have increased the scale of non-European

232 summary statistics and also varied the study dataset in multi-ethnic PRS studies[23,24,40].
233 These studies have shown that even when non-European cohorts are only a fraction the
234 size of the largest European study, they are likely to have disproportionate value for
235 predicting polygenic traits in other individuals of similar ancestry.
236
237 Given this background, we performed a systematic evaluation of polygenic prediction
238 accuracy across 17 quantitative anthropometric and blood panel traits and five disease
239 endpoints in British and Japanese individuals[23,57,58] by performing GWAS with the exact
240 same sample sizes in each population. We symmetrically demonstrate that prediction
241 accuracy is consistently higher with GWAS summary statistics from ancestry-matched
242 summary statistics (**Figure 4, Supplementary Figures 2-6**). Keeping in mind issues of
243 comparability described above, we note that BBJ is a hospital-based disease-
244 ascertained cohort, whereas UKBB is a healthier than average[59] population-based
245 cohort; thus, differences in observed heritability among these cohorts (rather than
246 among populations) due to differences in phenotype precision likely explain lower
247 prediction accuracy from the BBJ GWAS summary statistics for anthropometric and
248 blood panel traits, but higher prediction accuracy for five ascertained diseases
249 (**Supplementary Table 2**). Indeed, other East Asian studies have estimated higher
250 heritability for some quantitative traits than BBJ using the same methods, such as for
251 height (h^2 = 0.48±0.04 in Chinese women[60]). Some statistical fluctuations in the relative
252 differences in prediction accuracy across populations are likely driven by differences in
253 heritability measured in each population and/or trans-ethnic genetic correlation (i.e. of
254 common variant effect sizes at SNPs common in two populations, **Supplementary**

255 **Figures 7-10**, **Supplementary Tables 2–5**). These trans-ethnic correlation estimates
256 indicate that effect sizes were mostly highly correlated across ancestries, with a few
257 traits that were somewhat lower than excepted (e.g. height and BMI, with ρ_{ge}=0.69 and
258 0.75, respectively). Prediction accuracy was far lower in individuals of African descent in
259 the UK Biobank (**Supplementary Figures 4 and 11**) using GWAS summary statistics
260 from either European or Japanese ancestry individuals, consistent with reduced
261 prediction accuracy with increasing genetic divergence (**Figures 3 and 4**). These
262 population studies demonstrate the power and utility of increasingly diverse GWAS for
263 prediction, especially in populations of non-European descent.
264
265 While many other traits and diseases have been studied in multi-ethnic settings, few
266 have reported comparable metrics of prediction accuracy across populations.
267 Cardiovascular research, for example, has led the charge towards clinical translation of
268 PRS[1]. This enthusiasm is driven by observations that a polygenic burden of LDL-
269 increasing SNPs can confer monogenic-equivalent risk of cardiovascular disease, with
270 polygenic scores improving clinical models for risk assessment and statin prescription
271 that can reduce coronary heart disease and improve healthcare delivery efficiency[5-7].
272 However, many of these studies have been conducted exclusively in European descent
273 populations, with few studies rigorously evaluating population-level applicability to non-
274 Europeans. Those existing findings indeed demonstrate a large reduction in prediction
275 utility in non-European populations[11], though often with comparisons of odds ratios
276 among arbitrary breakpoints in the risk distribution that make comparisons across
277 studies challenging. To better clarify how polygenic prediction will be deployed in a

278 clinical setting with diverse populations, more systematic and thorough evaluations of
279 the utility of PRS within and across populations for many complex traits are still needed.
280 These evaluations would benefit from rigorous polygenic prediction accuracy
281 evaluations, especially for diverse non-European patients[61-63].
282
283 **Clinical use of PRS may uniquely exacerbate disparities**
284 Our impetus for raising these statistical issues limiting the generalizability of PRS across
285 population stems from our concerns that, while they are legitimately clinically promising
286 for improving health outcomes for many biomedical phenotypes, they may have a larger
287 potential to raise health disparities than other clinical factors for several reasons. The
288 opportunities they provide for improving health outcomes means they inevitably will and
289 should be pursued in the near term, but we urge that a concerted prioritization to make
290 GWAS summary statistics easily accessible for diverse populations and a variety of
291 traits and diseases is imperative, even when they are a fraction the size of the largest
292 existing European datasets. *Individual* clinical tests, biomarkers, and prescription drug
293 efficacy may vary across populations in their utility, but are fundamentally informed by
294 the same underlying biology[64,65]. Currently, guidelines state that as few as 120
295 individuals define reference intervals for clinical factors (though often smaller numbers
296 from only one subpopulation are used) and there is no clear definition of who is
297 "normal"[64]. Consequently, reference intervals for biomarkers can sometimes deviate
298 considerably by reported ethnicity[66-68]. Defining ethnicity-specific reference intervals is
299 clearly an important problem that can provide fundamental interpretability gains with
300 implications for some major health benefits (e.g. need for dialysis and development of

301 Type 2 diabetes based on ethnicity-specific serum creatinine and hemoglobin A1C
302 reference intervals, respectively)[67]. Simply put, some biomarkers or clinical tests scale
303 directly with health outcomes independent of ancestry, and many others may have
304 distributional differences by ancestry but are equally valid after centering with respect to
305 a readily collected population reference.
306
307 In contrast, PRS are uniformly less useful in understudied populations due to
308 differences in genomic variation and population history[13,14]. No analogous solution of
309 defining ethnicity-specific reference intervals would ameliorate health disparities
310 implications for PRS or fundamentally aid interpretability in non-European populations.
311 Rather, as we and others demonstrate, PRS are unique in that even with multi-ethnic
312 population references, these scores are fundamentally less informative in populations
313 more diverged from GWAS study cohorts.
314
315 The clinical use and deployment of genetic risk scores needs to be informed by the
316 issues surrounding tests that currently would unequivocally provide much greater
317 benefit to the subset of the world's population which is already on the positive end of
318 healthcare disparities. Conversely, African descent populations, which already endure
319 many of the largest health disparities globally, are often predicted marginally better, if at
320 all, compared to random (**Figure 4F**). They are therefore least likely to benefit from
321 improvements in precision healthcare delivery from genetic risk scores with existing
322 data due to human population history and study biases. This is a major concern globally
323 and especially in the U.S., which already leads other middle- and high-income countries

324 in both real and perceived healthcare disparities[69,70]. Thus, we would strongly urge that
325 any discourse on clinical use of PRS include a careful, quantitative assessment of the
326 economic and health disparities impacts on underrepresented populations that might be
327 unintentionally introduced, and raise awareness about how to eliminate these
328 disparities.
329
330 **How do we even the ledger?**
331 What can be done? The single most important step towards parity in PRS accuracy is
332 by vastly increasing the diversity of participants included and analyzed in genetic
333 studies, which will improve utility for all and most rapidly for underrepresented groups.
334 Regulatory protections against genetic discrimination are necessary to accompany calls
335 for more diverse studies; while some already exist in the U.S., including for health
336 insurance and employment opportunities via the Genetic Information Nondiscrimination
337 Act (GINA), stronger protections in these and other areas globally will be particularly
338 important for minorities and/or marginalized groups. An equal investment in GWAS
339 across all major ancestries and global populations is the most obvious solution to
340 generate a substrate for equally informative risk scores but is not likely to occur any
341 time soon absent a dramatic priority shift given the current imbalance and stalled
342 diversifying progress over the last five years (**Figure 1, Supplementary Figure 1**).
343 While it may be challenging or in some cases infeasible to acquire sample sizes large
344 enough for PRS to be equally informative in all populations, some much-needed efforts
345 towards increasing diversity in genomics that support open sharing of GWAS summary
346 data from multiple ancestries are underway. Examples include the *All of Us* Research

347 Program, the Population Architecture using Genomics and Epidemiology (PAGE)
348 Consortium, as well as some disease-focused consortia, such as the T2D-genes and
349 Stanley Global initiatives on the genetics of type II diabetes and psychiatric disorders,
350 respectively. Supporting data resources such as imputation panels, multi-ethnic
351 genotyping arrays, gene expression datasets from genetically diverse individuals, and
352 other tools are necessary to similarly empower these diverse studies for all populations.
353 The lack of supporting resources for diverse ancestries creates financial challenges for
354 association studies with limited resources, e.g. raising questions about whether to
355 genotype samples on GWAS arrays that may favor European allele frequencies versus
356 sequence samples, and how dense of an array to choose or how deeply to
357 sequence[71,72].
358
359 Additional leading global efforts also provide easy unified access linking genetic, clinical
360 record, and national registry data in more homogeneous continental ancestries, such as
361 the UK Biobank, BioBank Japan, China Kadoorie Biobank, and Nordic efforts (e.g. in
362 Danish, Estonian, Finnish, and other integrated biobanks). Notably, some of these
363 biobanks such as UK Biobank have participants with considerable global genetic
364 diversity that enables multi-ethnic comparisons; although minorities from this cohort
365 provide the largest deeply phenotyped GWAS cohorts for several ancestries, these
366 individuals are often excluded in current statistical analyses in favor of single ancestries,
367 large sample sizes, and the simplicity afforded by genetic homogeneity. These
368 considerations notwithstanding, there are critical needs and challenges for expanding
369 the scale of genetic studies of heritable traits in diverse populations; this is especially

370 apparent in Africa where humans originated and retain the most genetic diversity, as
371 Africans are understudied but disproportionately informative for genetic analyses and
372 evolutionary history[27,73]. The most notable investment here comes from the Human
373 Heredity and Health in Africa (H3Africa) Initiative, increasing genomics research
374 capacity in Africa through more than $216 million in funding from the NIH (USA) and
375 Wellcome Trust (UK) for genetics research led by African investigators[55,74]. The
376 increasing interest and scale of genetic studies in low- and middle-income countries
377 (LMICs) raises ethical and logistical considerations about data generation, access,
378 sharing, security, and analysis, as well as clinical implementation to ensure these
379 advances do not only benefit high-income countries. Frameworks such as the
380 H3ABioNet, a pan-African bioinformatics network designed to build capacity to enable
381 H3Africa researchers to analyze their data in Africa, provide cost-effective examples for
382 training local scientists in LMICs[75].
383
384 The prerequisite data for dramatically increasing diversity also hypothetically exist in
385 several large-scale publicly funded datasets such as the Million Veterans Project and
386 Trans-Omics for Precision Medicine (TOPMed), but with problematic data access issues
387 in which even GWAS summary data within and across populations are not publicly
388 shared. Existing GWAS consortia also need to carefully consider the granularity of
389 summary statistics they release, as finer scale continental ancestries and phenotypes in
390 large, multi-ethnic projects enable ancestry-matched analyses not possible with a single
391 set of summary statistics. While there is an understandable patient privacy balance to
392 strike when sharing individual-level data, GWAS summary statistics from all publicly

393 funded and as many privately funded projects as possible should be made easily and
394 publicly accessible to improve global health outcomes. Efforts to unify phenotype
395 definitions, normalization approaches, and GWAS methods among studies will also
396 improve comparability.
397
398 To enable progress towards parity, it will be critical that open data sharing standards be
399 adopted for all ancestries and for genetic studies of all sample sizes, not just the largest
400 European results. Locally appropriate and secure genetic data sharing techniques as
401 well as equitable technology availability will need to be adopted widely in Asia and
402 Africa as they are in Europe and North America, to ensure that maximum value is
403 achieved from existing and ongoing efforts that are being developed to help counter the
404 current imbalance. Simultaneously, ethical considerations require that research capacity
405 is increased in LMICs with simultaneous growth of diverse population studies to balance
406 the benefits of these studies to scientists and patients globally versus locally to ensure
407 that everyone benefits. Methodological improvements that better define risk scores by
408 accounting for population allele frequency, LD, and/or admixture differences
409 appropriately are underway and may help considerably but will not by themselves bring
410 equality. All of these efforts are important and should be prioritized not just for risk
411 prediction but more generally to maximize the use and applicability of genetics to inform
412 on the biology of disease. Given the acute recent attention on clinical use of PRS, we
413 believe it is paramount to recognize their potential to improve health outcomes for all
414 individuals and many complex diseases. Simultaneously, we as a field must address the
415 disparity in utility in an ethically thoughtful and scientifically rigorous fashion, lest we

416 inadvertently enable genetic technologies to contribute to, rather than reduce, existing
417 health disparities.
418

419 **Author contributions**

420 A.R.M. and M.J.D. conceived and designed the experiments. A.R.M. and M.K.
421 performed statistical analysis. A.R.M. and M.K. analyzed the data. A.R.M., M.K., Y.K.,
422 Y.O., B.M.N., and M.J.D. contributed reagents/materials/analysis tools. A.R.M., M.K.,
423 B.M.N., and M.J.D. wrote the paper.
424

425 **Competing interests**

426 The authors declare no competing interests.
427

428 **Acknowledgments**

429 We thank Amit Khera for helpful discussions. We also thank Michiaki Kubo, Yoshinori
430 Murakami, Masato Akiyama, and Kazuyoshi Ishigaki for their support in the BioBank
431 Japan Project analysis. We are grateful to Steven Gazal for his help in calculating LD
432 scores. This work was supported by funding from the National Institutes of Health
433 (K99MH117229 to A.R.M.). UK Biobank analyses were conducted via application
434 31063. The BioBank Japan Project was supported by the Tailor-Made Medical
435 Treatment Program of the Ministry of Education, Culture, Sports, Science, and
436 Technology (MEXT) and the Japan Agency for Medical Research and Development
437 (AMED). M.K. was supported by a Nakajima Foundation Fellowship and the Masason
438 Foundation.

Figures

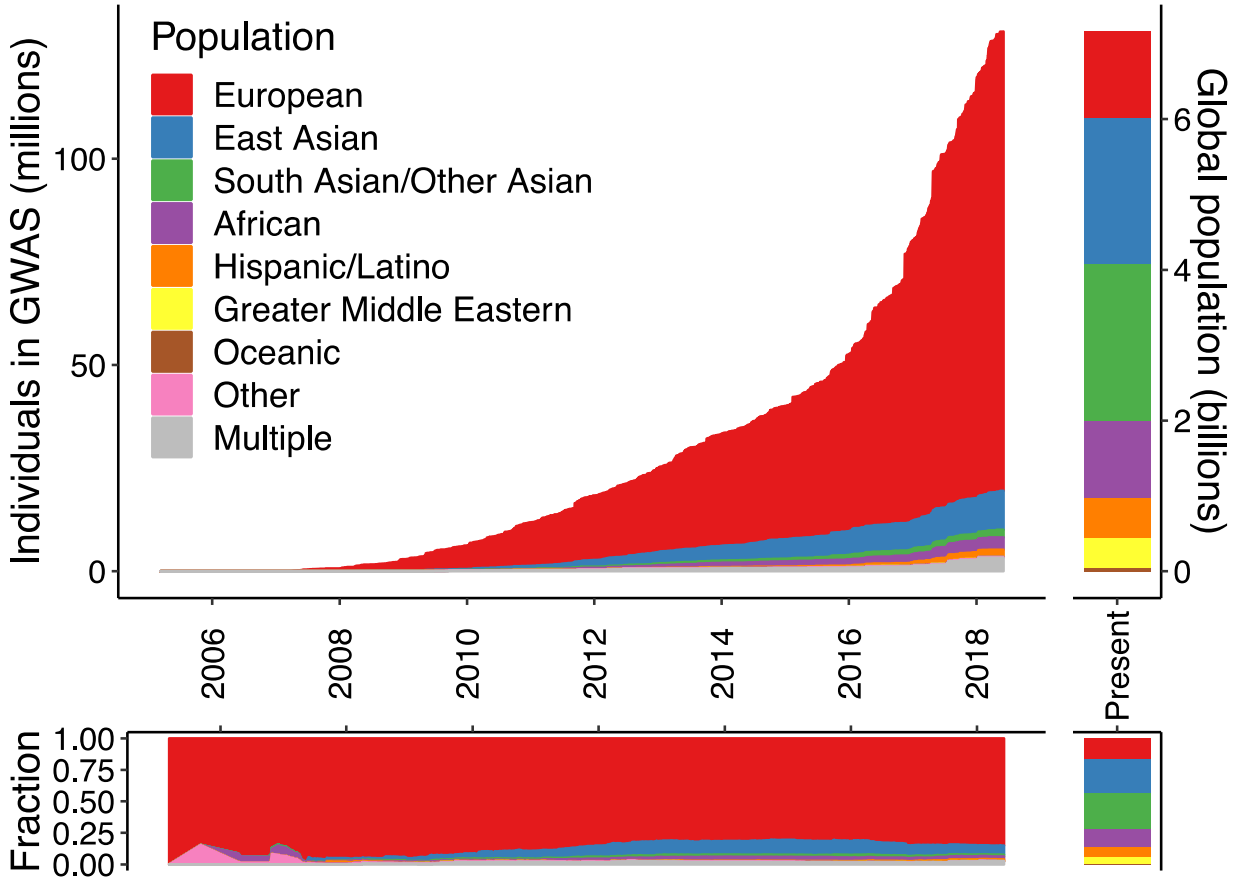

Figure 1 – Ancestry of GWAS participants over time compared to the global population. Cumulative data as reported by the GWAS catalog[76]. Individuals whose ancestry is "not reported" are not shown.

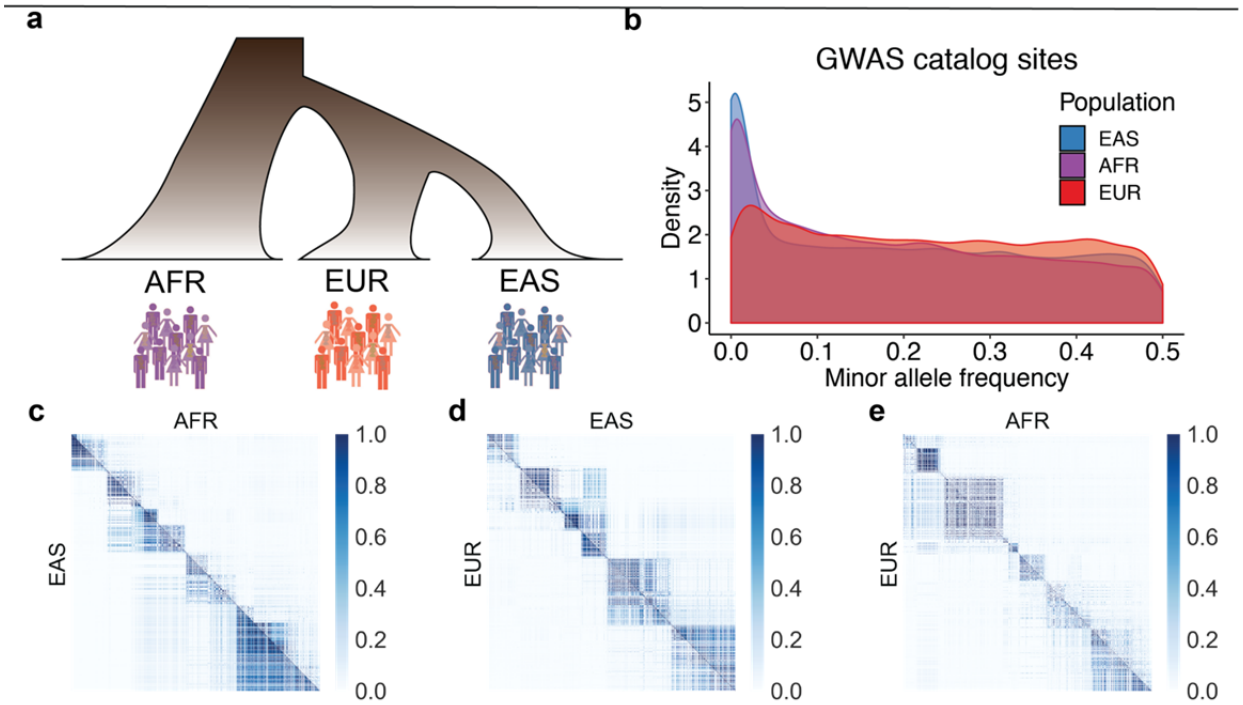

Figure 2 – Demographic relationships, allele frequency differences, and local LD patterns between population pairs. Data analyzed from 1000 Genomes, in which population labels are: AFR = continental African, EUR = European, and EAS = East Asian. **a**) Cartoon relationships among AFR, EUR, and EAS populations. **b**) Allele frequency distributions in AFR, EUR, and EAS populations of variants from the GWAS catalog. **c-e**) Color axis shows LD scale (r^2). LD comparisons between pairs of populations show the same region of the genome for each comparison (representative region is chr1, 51572kb-52857kb) among pairs of SNPs polymorphic in both populations, illustrating that different SNPs are polymorphic across some population pairs, and that these SNPs have variable LD patterns across populations.

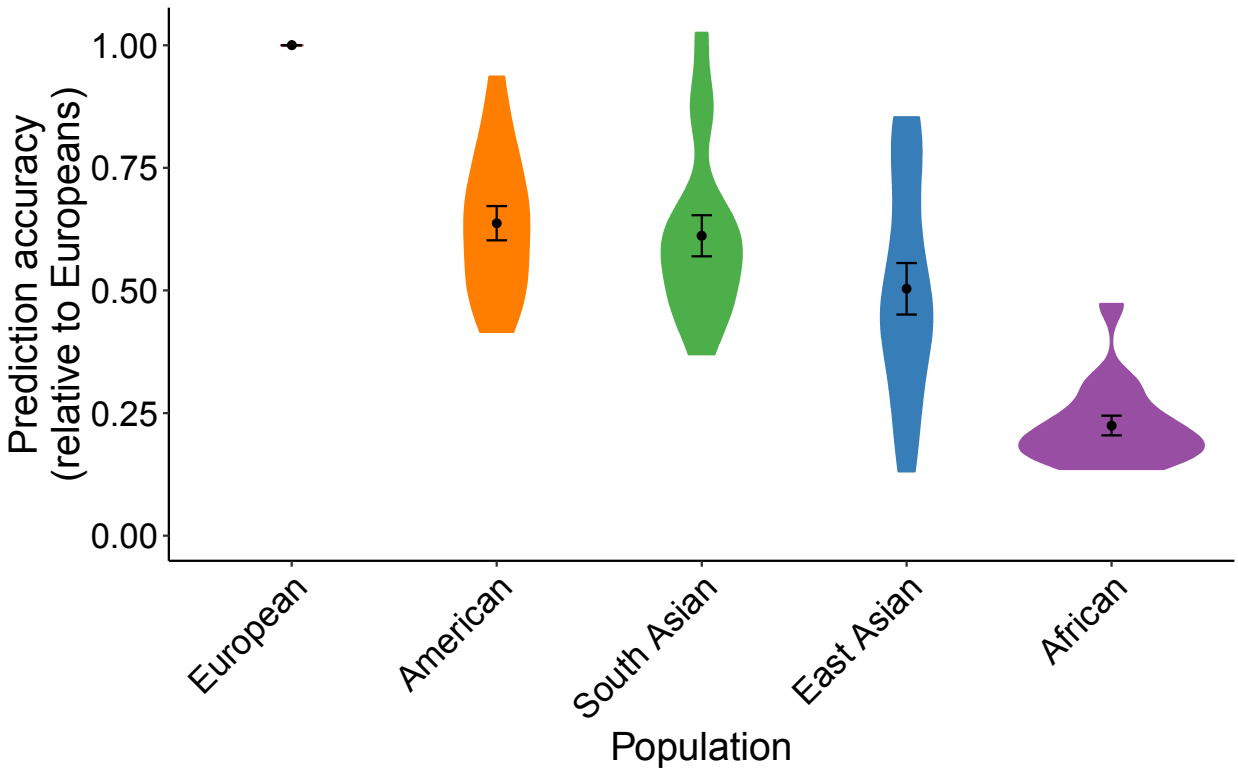

Figure 3 – Prediction accuracy relative to European ancestry individuals across 17 quantitative traits and 5 continental populations in UKBB. All phenotypes shown here are quantitative anthropometric and blood panel traits, as described in **Supplementary Table 6**, which includes discovery cohort sample sizes. Prediction target individuals do not overlap with the discovery cohort and are unrelated, with sample sizes shown in **Supplementary Table 7**. Violin plots show distributions of relative prediction accuracies, points show mean values, and error bars show standard errors of the means. Prediction R^2 for each trait and population are shown in **Supplementary Figure 12**.

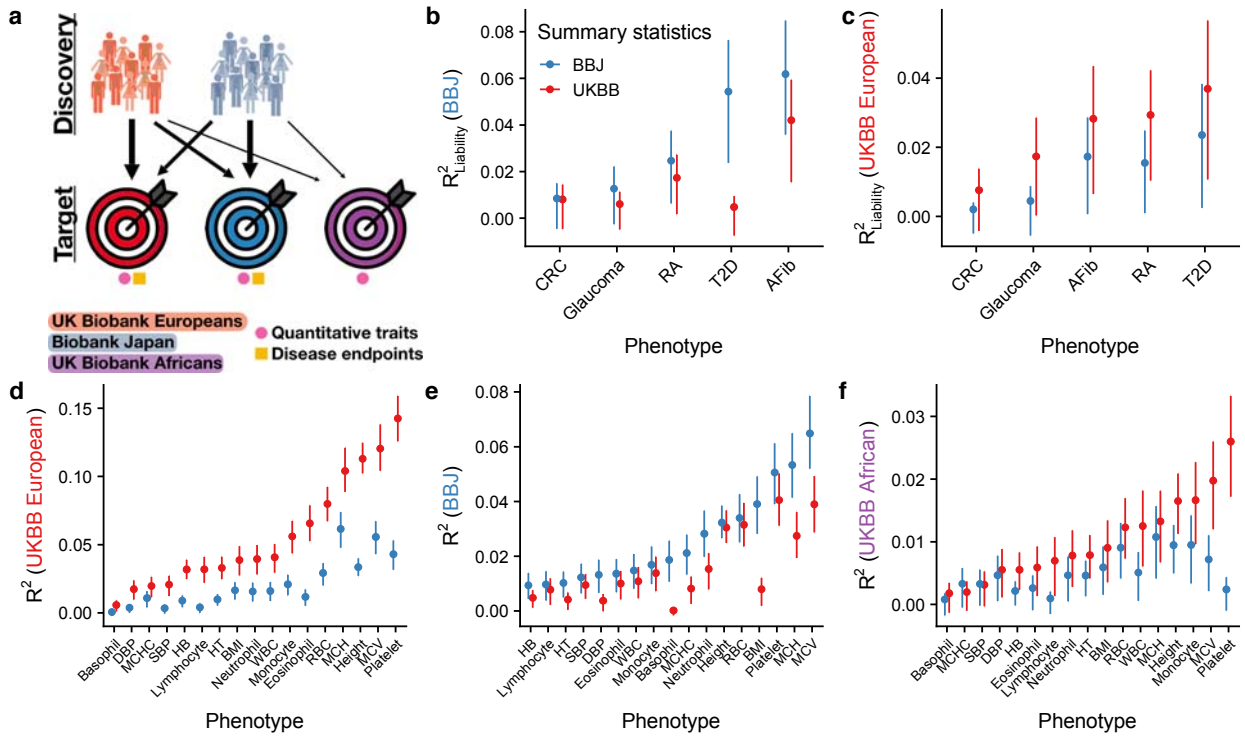

Figure 4 – Polygenic risk prediction accuracy in Japanese, British, and African descent individuals using independent GWAS of equal sample sizes in the BioBank Japan (BBJ) and UK Biobank (UKBB). **a**) Explanatory diagram showing the different discovery and target cohorts/populations, and disease endpoints versus quantitative traits. **b-f**) Genetic prediction accuracy computed from independent BBJ and UKBB summary statistics with identical sample sizes (**Supplementary Tables 6 and 8**). Note that y-axes differ, reflecting differences in prediction accuracy. **b-c**) PRS accuracy for five diseases in: Japanese individuals in the BBJ (**b**) and British individuals in the UKBB. **d-f**) PRS accuracy for 17 anthropometric and blood panel traits in: Japanese individuals in the BBJ (**d**), British individuals in the UKBB (**e**), and African descent British individuals in the UKBB (**f**). Trait abbreviations are as in **Supplementary Table 6**. Each point shows the maximum R^2 (i.e. best predictor) across five p-value thresholds, and lines correspond to 95% confidence intervals calculated via bootstrap.

482 R^2 values for all p-value thresholds tested are shown in **Supplementary Figures 2-6**.
483 Prediction accuracy tends to be higher in the UKBB for quantitative traits than in BBJ
484 and vice versa for disease endpoints, likely because of concomitant phenotype
485 precision and consequently observed heritability for these classes of traits
486 (**Supplementary Tables 2-4**). Thalassemia and sickle cell disease are unlikely to
487 explain a significant fraction of prediction accuracy differences for blood panels across
488 populations, as few individuals have been diagnosed with these disorders via ICD-10
489 codes (**Supplementary Table 9**).

References

1. Knowles, J. W. & Ashley, E. A. Cardiovascular disease: The rise of the genetic risk score. *PLoS Med* **15,** e1002546–7 (2018).
2. Maas, P. *et al.* Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol* **2,** 1295–8 (2016).
3. Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* **50,** 928–936 (2018).
4. Sharp, S. A. *et al.* Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis. *Diabetes Care* **42,** 200–207 (2019).
5. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50,** 1219–1224 (2018).
6. Kullo, I. J. *et al.* Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). *Circulation* **133,** 1181–1188 (2016).
7. Natarajan, P. *et al.* Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. *Circulation* **135,** 2091–2101 (2017).
8. Paquette, M. *et al.* Polygenic risk score predicts prevalence of cardiovascular disease in patients with familial hypercholesterolemia. *Journal of Clinical Lipidology* **11,** 725–732.e5 (2017).
9. Tikkanen, E., Havulinna, A. S., Palotie, A., Salomaa, V. & Ripatti, S. Genetic Risk Prediction and a 2-Stage Risk Screening Strategy for Coronary Heart Disease. *Arterioscler. Thromb. Vasc. Biol.* **33,** 2261–2266 (2013).
10. Frieser, M. J., Wilson, S. & Vrieze, S. Behavioral impact of return of genetic test results for complex disease: Systematic review and meta-analysis. *Health Psychol* **37,** 1134–1144 (2018).
11. Khera, A. V. *et al.* Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med* **375,** 2349–2358 (2016).
12. Khera, A. V. & Kathiresan, S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nature Publishing Group* **18,** 331–344 (2017).
13. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100,** 635–649 (2017).
14. Scutari, M., Mackay, I. & Balding, D. Using Genetic Distance to Infer the Accuracy of Genomic Prediction. *PLoS Genet* **12,** e1006288–19 (2016).
15. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97,** 576–592 (2015).
16. Ware, E. B. *et al.* Heterogeneity in polygenic scores for common human traits. *biorxiv.org*
doi:10.1101/106062
17. Curtis, D. Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *Psychiatric Genetics* **28,** 85–89 (2018).

534 18. Schizophrenia Working Group of the Psychiatric Genomics Consortium.
535 Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,**
536 421–427 (2014).
537 19. Belsky, D. W. *et al.* Development and Evaluation of a Genetic Risk Score for
538 Obesity. *Biodemography and Social Biology* **59,** 85–100 (2013).
539 20. Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M. & Boardman, J. D.
540 Polygenic Influence on Educational Attainment. *AERA Open* **1,**
541 233285841559997–13 (2015).
542 21. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide
543 association study of educational attainment in 1.1 million individuals. *Nature*
544 *Publishing Group* 1–16 (2018). doi:10.1038/s41588-018-0147-3
545 22. Vassos, E. *et al.* An Examination of Polygenic Score Risk Prediction in
546 Individuals With First-Episode Psychosis. *Biological Psychiatry* **81,** 470–477
547 (2017).
548 23. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for
549 body mass index in the Japanese population. *Nat Genet* **49,** 1458–1467 (2017).
550 24. Li, Z. *et al.* Genome-wide association analysis identifies 30 new susceptibility loci
551 for schizophrenia. *Nat Genet* **49,** 1576–1583 (2017).
552 25. Need, A. C. & Goldstein, D. B. Next generation disparities in human genomics:
553 concerns and remedies. *Trends in Genetics* **25,** 489–494 (2009).
554 26. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538,**
555 161–164 (2016).
556 27. Morales, J. *et al.* A standardized framework for representation of ancestry data in
557 genomics studies, with application to the NHGRI-EBI GWAS Catalog. 1–10
558 (2018). doi:10.1186/s13059-018-1396-2
559 28. Rosenberg, N. A. *et al.* Genome-wide association studies in diverse populations.
560 *Nature Publishing Group* **11,** 356–366 (2010).
561 29. Sham, P. C., Cherny, S. S., Purcell, S. & Hewitt, J. K. Power of linkage versus
562 association analysis of quantitative traits, by use of variance-components
563 models, for sibship data. *The American Journal of Human Genetics* **66,** 1616–
564 1630 (2000).
565 30. The 1000 Genomes Project Consortium *et al.* A global reference for human
566 genetic variation. *Nature* **526,** 68–74 (2015).
567 31. Consortium, T. S. 2. D. Sequence variants in SLC16A11 are a common risk
568 factor for type 2 diabetes in Mexico. *Nature* **506,** 97–101 (2014).
569 32. Estrada, K. *et al.* Association of a Low-Frequency Variant in HNF1AWith Type 2
570 Diabetes in a Latino Population. *JAMA* **311,** 2305–10 (2014).
571 33. Haiman, C. A. *et al.* Genome-wide association study of prostate cancer in men of
572 African ancestry identifies a susceptibility locus at 17q21. *Nature Publishing*
573 *Group* **43,** 570–573 (2011).
574 34. Genovese, G. *et al.* Association of trypanolytic ApoL1 variants with kidney
575 disease in African Americans. *Science* **329,** 841–845 (2010).
576 35. International Multiple Sclerosis Genetics Consortium *et al.* Association analyses
577 identify 38 susceptibility loci for inflammatory bowel disease and highlight shared
578 genetic risk across populations. *Nat Genet* **47,** 979–986 (2015).

36. Carlson, C. S. *et al.* Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study. *PLoS Biol* **11,** e1001661–11 (2013).
37. Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447,** 1087–1093 (2007).
38. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* **46,** 234–244 (2014).
39. Waters, K. M. *et al.* Consistent Association of Type 2 Diabetes Risk Variants Found in Europeans in Diverse Racial and Ethnic Groups. *PLoS Genet* **6,** e1001078–9 (2010).
40. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. 1–41 (2018). doi:10.1101/445874
41. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48,** 1279–1283 (2016).
42. Huang, H. *et al.* Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* 1–19 (2017). doi:10.1038/nature22969
43. Sohail, M. *et al.* Signals of polygenic adaptation on height have been overestimated due to uncorrected population structure in genome-wide association studies. 1–44 (2018). doi:10.1101/355057
44. Berg, J. J. *et al.* Reduced signal for polygenic adaptation of height in UK Biobank. 1–44 (2018). doi:10.1101/354951
45. Kerminen, S. *et al.* Geographic variation and bias in polygenic scores of complex diseases and traits in Finland:. 1–35 (2018). doi:10.1101/485441
46. Novembre, J. & Barton, N. H. Tread Lightly Interpreting Polygenic Tests of Selection. *Genetics* **208,** 1351–1355 (2018).
47. Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nature Publishing Group* **16,** 333–343 (2015).
48. Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am. J. Hum. Genet.* **99,** 76–88 (2016).
49. Galinsky, K. J. *et al.* Estimating cross-population genetic correlations of causal effect sizes. *Genet. Epidemiol.* **200,** 1285 (2018).
50. Li, D., Zhao, H. & Gelernter, J. Strong protective effect of the aldehyde dehydrogenase gene (ALDH2) 504lys (*2) allele against alcoholism and alcohol-induced medical diseases in Asians. *Human Genetics* **131,** 725–737 (2011).
51. Zhu, Z. *et al.* Dominance Genetic Variation Contributes Little to the Missing Heritability for Human Complex Traits. *The American Journal of Human Genetics* **96,** 377–385 (2015).
52. Paré, G., Mao, S. & Deng, W. Q. A machine-learning heuristic to improve gene score prediction of polygenic traits. *Sci Rep* **7,** 12665 (2017).
53. Martin, A. R. *et al.* An Unexpectedly Complex Architecture for Skin Pigmentation in Africans. *Cell* **171,** 1340–1353.e14 (2017).

623 54. Duncan, L. E. *et al.* Largest GWAS of PTSD (N=20 070) yields genetic overlap
624 with schizophrenia and sex differences in heritability. *Molecular Psychiatry* **23,**
625 666–673 (2018).
626 55. H3Africa Consortium *et al.* Research capacity. Enabling the genomic revolution
627 in Africa. *Science* **344,** 1346–1348 (2014).
628 56. Hindorff, L. A. *et al.* Prioritizing diversity in human genomics research. *Nature*
629 *Publishing Group* **19,** 175–185 (2018).
630 57. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population
631 links cell types to complex human diseases. *Nat Genet* **50,** 390–400 (2018).
632 58. Howrigan, D. Details and Considerations of the UK Biobank GWAS. (2017).
633 Available at: http://scholar.google.comjavascript:void(0). (Accessed: 9 November
634 2017)
635 59. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related
636 Characteristics of UK Biobank Participants With Those of the General
637 Population. *Am. J. Epidemiol.* **186,** 1026–1034 (2017).
638 60. Liu, S. *et al.* Genomic Analyses from Non-invasive Prenatal Testing Reveal
639 Genetic Associations, Patterns of Viral Infections, and Chinese Population
640 History. *Cell* **175,** 347–359.e14 (2018).
641 61. Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nature*
642 *Publishing Group* **14,** 507–515 (2013).
643 62. Wray, N. R. *et al.* Research Review: Polygenic methods and their application to
644 psychiatric traits. *J Child Psychol Psychiatr* **55,** 1068–1087 (2014).
645 63. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of
646 polygenic risk scores. *Nature Publishing Group* **19,** 581–590 (2018).
647 64. Manrai, A. K., Patel, C. J. & Ioannidis, J. P. A. In the Era of Precision Medicine
648 and Big Data, Who Is Normal? *JAMA* **319,** 1981–3 (2018).
649 65. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets
650 through human genetics. *Nature Publishing Group* **12,** 581–594 (2013).
651 66. Carroll, M. D., Kit, B. K., Lacher, D. A., Shero, S. T. & Mussolino, M. E. Trends in
652 Lipids and Lipoproteins in US Adults, 1988-2010. *JAMA* **308,** 1545–1554 (2012).
653 67. Rappoport, N. *et al.* Comparing Ethnicity-Specific Reference Intervals for Clinical
654 Laboratory Tests from EHR Data. *Jrnl App Lab Med* **3,** 366–377 (2018).
655 68. Lim, E., Miyamura, J. & Chen, J. J. Racial/Ethnic-Specific Reference Intervals for
656 Common Laboratory Tests: A Comparison among Asians, Blacks, Hispanics, and
657 White. *Hawaii J Med Public Health* **74,** 302–310 (2015).
658 69. Hero, J. O., Zaslavsky, A. M. & Blendon, R. J. The United States Leads Other
659 Nations In Differences By Income In Perceptions Of Health And Health Care.
660 *Health Affairs* **36,** 1032–1040 (2017).
661 70. Williams, D. R., Priest, N. & Anderson, N. B. Understanding associations among
662 race, socioeconomic status, and health: Patterns and prospects. *Health*
663 *Psychology* **35,** 407–411 (2016).
664 71. Gilly, A. *et al.* Very low depth whole genome sequencing in complex trait
665 association studies. *Bioinformatics* (2018). doi:10.1093/bioinformatics/bty1032
666 72. Pasaniuc, B. *et al.* Extremely low-coverage sequencing and imputation increases
667 power for genome-wide association studies. *Nat Genet* **44,** 631–635 (2012).

73. Martin, A. R., Teferra, S., Möller, M., Hoal, E. G. & Daly, M. J. The critical needs and challenges for genetic architecture studies in Africa. *Current Opinion in Genetics & Development* **53,** 113–120 (2018).
74. Coles, E. & Mensah, G. A. Geography of Genetics and Genomics Research Funding in Africa. *Global Heart* **12,** 173–176 (2017).
75. Mulder, N. J. *et al.* Development of Bioinformatics Infrastructure for Genomics Research. *Global Heart* **12,** 91–98 (2017).
76. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucl. Acids Res.* **45,** D896–D901 (2017).