Hidden 'risk' in polygenic scores: clinical use today could exacerbate 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

Abstract

Polygenic risk scores have the potential to improve health outcomes for a variety of complex diseases and are poised for clinical translation, driven by the low cost of genotyping (<$50 per person), the ability to predict genetic risk of many diseases with a single test, and the dramatically increasing scale and power of genetic studies that aid prediction accuracy. However, the major *ethical* and *scientific* challenge surrounding clinical implementation is the observation that they are currently of far greater predictive value in individuals of recent European ancestry than others. The better performance of such risk scores in European populations is an inescapable consequence of the heavily biased makeup of genome-wide association studies, with an estimated 79% of participants in all existing genetic studies being of European descent. Empirically, polygenic risk scores perform far better in European populations, with prediction accuracy reduced by approximately 2- to 5-fold in East Asian and African American populations, respectively. This highlights that—unlike specific 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 prediction today would systematically afford greater improvement to European populations. Early diversifying efforts, however, show promise in levelling this vast imbalance, even when non-European sample sizes are considerably smaller than the best-powered studies to

date. Polygenic risk scores provide a new opportunity to improve health outcomes for many diseases in all populations, but to realize this full potential equitably, we must prioritize greater inclusivity of diverse study participants in genetic studies and open access to resulting 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 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 criticisms of limited prediction power are now recognized to have been chiefly an issue of small sample size, which is no longer the case for many outcomes [1]. For example, integrated models of PRS together with other lifestyle and clinical factors have enabled clinicians to more accurately quantify the risk of heart attack for patients; consequently, they have more effectively targeted the reduction of LDL cholesterol and by extension heart attack by prescribing statins to patients at the greatest overall risk of cardiovascular disease [2-6]. While we share enthusiasm about the potential of PRS to improve health outcomes through their eventual routine implementation as clinical biomarkers, we consider the consistent observation that they are currently of far greater predictive value in individuals of recent European descent than in others to be the major *ethical* and

69 *scientific* challenge surrounding clinical translation and, at present, the most critical
70 limitation to genetics in precision medicine. The scientific basis of this imbalance has
71 been demonstrated in population genetics simulations, theoretically, and empirically
72 across many traits and diseases [7-18].
73
74 All studies to date using well-powered genome-wide association studies (GWAS) to
75 assess the predictive value of PRS in European and non-European descent populations
76 have made a consistent observation: PRS predict individual risk far better in Europeans
77 than non-Europeans. In complex traits including height, body mass index (BMI),
78 educational attainment, schizophrenia, and major depression, existing PRS computed
79 with the largest available GWAS results predict outcomes far more accurately in new
80 samples of European-descent than they do in non-Europeans, with the clearest study
81 examples in East Asians and African Americans [11,12,14-20]. Rather than chance or
82 biology, this is a predictable consequence of the fact that the genetic discovery efforts to
83 date heavily overrepresent European populations globally. The correlation between true
84 and genetically predicted phenotypes decays with genetic divergence from the makeup
85 of the discovery GWAS, meaning that the accuracy of polygenic scores in different
86 populations is highly dependent on the study population representation in the largest
87 existing 'training' GWAS. Here, we document study biases that underrepresent non-
88 European populations in current GWAS, and explain the fundamental concepts
89 contributing to reduced phenotypic variance explained with increasing genetic
90 divergence from populations included in GWAS.
91

92 **Predictable basis of disparities in polygenic risk score accuracy**

93 The lack of generalizability of genetic studies across global populations arises from the
94 overwhelming abundance of European descent studies—according to the GWAS
95 catalog [21-24], ~79% of all GWAS participants are of European descent despite making
96 up only 16% of the global population (**Figure 1**). More concerningly, the fraction of non-
97 European individuals in GWAS has stagnated or declined since late 2014 (**Figure 1**),
98 suggesting that we are not on a trajectory to correct this imbalance. These numbers
99 provide a composite metric of study availability, accessibility, and use—i.e., cohorts that
100 have been included in numerous studies are represented multiple times, which may
101 disproportionately include cohorts of European descent. The relative sample
102 compositions of GWAS result in highly predictable disparities in prediction accuracy;
103 statistical and population genetics theory predicts that genetic risk prediction accuracy
104 will decay with increasing genetic divergence between the original GWAS sample and
105 target of prediction, a function of population history [9,10]. This pattern can be attributed to
106 several statistical observations which we detail below: 1) GWAS favor the discovery of
107 genetic variants that are common in the study population; 2) linkage disequilibrium (LD)
108 differentiates marginal effect size estimates for highly polygenic traits across
109 populations, even when causal variants are the same; and 3) demographic and
110 environmental differences may drive differential forces of natural selection that in turn
111 drive differences in causal genetic architecture. Of note, the first two of these degrade
112 prediction performance across populations substantially even when there exist no
113 biological, environmental, or diagnostic differences.
114

Common discoveries and low-hanging fruit

First, the power to discover an association in a genetic study depends on the effect size and frequency of the variant [25]. This power dependence means that the most significant associations tend to be more common in the populations in which they are discovered than in other populations [9,26]. For example, GWAS catalog variants are on average more common in European populations compared to East Asian and African populations (**Figure 2B**), an observation not representative of genomic variants at large. Understudied populations offer low-hanging fruit for genetic discovery because variants that are common in these groups but rare or absent in European populations could not be discovered even with very large European sample sizes. Some examples include *SLC16A11* and *HNF1A* associations with type II diabetes in Latino populations, *APOL1* associations with end-stage kidney disease, and associations with prostate cancer in African descent populations [27-30]. If we assume that causal genetic variants have an equal effect across all populations—an assumption with some empirical support that offers the best case scenario for transferability [31-35]—Eurocentric GWAS biases mean that variants that are common in European populations are preferentially discovered and associated with risk, and thus account for a larger fraction of the variance in polygenic risk [9]. Furthermore, imputation reference panels share the same biases as in GWAS, and imputing sites that are common in European populations but rarer in other populations is challenging when the catalog of non-European haplotypes is substantially smaller. These issues are insurmountable through statistical methods alone, but rather motivate substantial investments in more diverse populations to produce similar-sized

GWAS of biomedical phenotypes as well as sequenced reference panels in other populations.

Linkage disequilibrium

Second, the correlation structure of the human genome, i.e. LD, varies across populations due to demographic history (**Figure 2A,C-E**). These LD differences in turn drive differences in effect size estimates (i.e. predictors) from GWAS across populations, even when causal effects are the same. (Mathematically, the marginal GWAS estimate $\hat{\beta}_j = \sum_{k=1}^{m} r_{j,k}\beta_k + \epsilon_j$, where $\hat{\beta}_j$ are effect size estimates at SNP *j*, $r_{j,k}$ is pairwise SNP LD between SNPs *j* and *k*, β_k is the causal SNP effect at nearby SNP *k*, and ϵ is residual error from bias or noise). While differences in effect size estimates due to LD differences may typically be small for most regions of the genome, PRS sum across these effects, also aggregating these population differences. Statistical methods that account for LD differences across populations may help improve risk prediction accuracy within each population. While empirical studies suggest that causal effect sizes tend to be shared [31,32], it may not be feasible to fine-map most variants to a single locus to solve issues of low generalizability, even with very large GWAS (i.e., millions). This is because complex traits are highly polygenic, meaning most of our prediction power comes from small effects that do not meet genome-wide significance and/or cannot be fine-mapped, even in the best-powered GWAS to date [36].

History, selection, the environment, and complex interactions

Lastly, other environmental, demographic, and cohort considerations may further worsen prediction accuracy differences across populations in less predictable ways. GWAS ancestry study biases and LD differences across populations are extremely challenging to address, but these issues actually make many favorable assumptions that all causal loci have the same impact and are under equivalent selective pressure in all populations. In contrast, other effects on polygenic adaptation or risk scores such as natural selection can impact populations differently based on their unique histories. Additionally, residual uncorrected population stratification may impact risk prediction accuracy across populations, but the magnitude of its effect is currently unclear. These effects are particularly challenging to disentangle, as has clearly been demonstrated for height, where evidence of polygenic adaptation is under question [37,38]. Comparisons of geographically stratified phenotypes like height across populations with highly divergent genetic backgrounds and mean environmental differences, such as differences in resource abundance during development across continents, are especially prone to uninterpretable results [39]. Related to stratification, most polygenic scoring methods do not explicitly address recent admixture and none consider recently admixed individuals' unique local mosaic of ancestry—further methods development in this space is needed. Furthermore, comparing PRS across environmentally stratified cohorts, such as in some biobanks with healthy volunteer effects versus disease study datasets or hospital-based cohorts, requires careful consideration of technical differences, collider bias, as well as variability in baseline health status among studies. It is also important to consider differences in clinical definition of the phenotypes and heterogeneous constitution of sub-phenotypes among countries.

Differences in environmental exposure, gene × environment interactions, historical population size dynamics, and other factors will further limit generalizability for genetic risk scores in an unpredictable, trait-specific fashion [40,41]. While non-linear genetic factors explain little variation in complex traits beyond a purely additive model [42], some unrecognized nonlinearities and gene × gene interactions can also induce genetic risk prediction challenges, as pairwise interactions are likely to vary more across populations than individual SNPs. Mathematically, we can simplistically think of this in terms of a two-SNP model, in which the sum of two SNP effects is likely to explain more phenotypic variance than the product of the same SNPs. Some machine learning approaches may thus modestly improve genetic prediction accuracy for some phenotypes [43], but these approaches are most likely to improve prediction accuracy for atypical traits with simpler architectures, known interactions, and poor prediction generalizability across populations, such as skin pigmentation [44].

Limited generalizability of genetic prediction across diverse populations

Previous work has assessed prediction accuracy across diverse populations in several traits and diseases for which GWAS summary statistics are available. These assessments are becoming increasingly feasible with the growth and public availability of global biobanks for quantitative traits as well as diversifying priorities from funding agencies [45,46]. As of yet, multi-ethnic work has been slow in most disease areas [47], limiting even the opportunity to assess prediction utility in non-European cohorts. Nonetheless, we have assembled prediction accuracy statistics from several studies

using the largest European GWAS to predict several phenotypes in target European and non-European cohorts. For example, multiple schizophrenia studies consistently predicted risk on average 2.2-fold worse in East Asians relative to Europeans, (i.e. μ=0.46, σ=0.06), using summary statistics from a Eurocentric GWAS [11,14] (**Figure 3**), despite the fact that there is no genetic heterogeneity in schizophrenia between the two populations. This finding is even more pronounced in African Americans, where genetic divergence from Europeans is greater than between Europeans and East Asians [26]. Across several phenotypes with a range of genetic architectures in which empirical evaluations were available, including BMI, educational attainment, height, and schizophrenia, prediction accuracy using European GWAS summary statistics was on average 4.5-fold less accurate in African Americans than in Europeans (i.e. μ=0.22, σ=0.09, **Figure 3**) [11,12,15-18]. By extension, prediction accuracy is expected to be even lower in African Americans with higher than average African ancestry or among populations with greater divergence from Europeans (e.g. some southern African populations). These enormous disparities are not simply methodological issues, as various approaches (e.g. pruning and threshold versus LDPred) and accuracy metrics (R^2 for quantitative traits and various pseudo-R^2 metrics for binary traits) illustrate this consistently poorer performance in populations distinct from the discovery sample across a range of polygenic traits (**Table S2**).

Prioritizing diversity shows early promise for polygenic prediction

Early diversifying GWAS efforts have been especially productive for informing on these questions surrounding risk prediction. Rather than varying the prediction target dataset,

some GWAS in diverse populations have increased the scale of non-European summary statistics and also varied the study dataset in multi-ethnic PRS studies. For example, a BioBank Japan GWAS study (N=158,284) showed that compared to a 2× larger European GWAS (N=322,154), the variance in BMI explained in an independent Japanese cohort with Japanese GWAS summary statistics was on average 1.5-fold greater than with European GWAS summary statistics (R^2=0.154 vs 0.104 at $p < 0.05$, respectively) [19]. Similarly, a Chinese schizophrenia study (N=7,699 cases and 18,327 controls) showed that compared to an effectively 5-fold larger European GWAS (N=36,989 cases, 113,075 controls), prediction accuracy in an independent Chinese cohort with GWAS summary statistics from China far surpassed prediction accuracy from European summary statistics by 2.6-fold (2.3% versus 6.2%) [20]. Thus, even when studies in non-European populations are only a fraction the size of the largest European study, they are likely to have disproportionate value for predicting polygenic traits in other individuals of similar ancestry.

Given this background, we performed a systematic evaluation of polygenic prediction accuracy across 17 quantitative anthropometric and blood panel traits in British and Japanese individuals [19,48,49] by performing GWAS with the exact same sample sizes in each population. We symmetrically demonstrate that prediction accuracy is consistently higher with GWAS summary statistics from ancestry-matched summary statistics (**Figure 4**). Keeping in mind issues of comparability described above, we note that the BioBank Japan (BBJ) is a hospital-based cohort, whereas UK Biobank (UKBB) is a healthier than average population-based cohort, and that differences in observed

heritability among these cohorts (rather than among populations) likely explain lower prediction accuracy from the BBJ GWAS summary statistics (**Table S3**). Some statistical fluctuations in the relative differences in prediction accuracy across populations are likely driven by differences in trans-ethnic genetic correlation (i.e. comparing across ancestries the estimated correlation of common variant effect sizes at SNPs common in both populations via Popcorn) and/or differences in heritability measured in each population (**Figure S1**, **Table S3**). Prediction accuracy was far lower in individuals of African descent in the UK Biobank (**Figure S5**) using GWAS summary statistics from European or Japanese ancestry individuals (**Figure 4**). These population studies demonstrate the power and utility of increasingly diverse GWAS for prediction, especially in populations of non-European descent.

While many other traits and diseases have been studied in multi-ethnic settings, few have reported comparable metrics of prediction accuracy across populations. Cardiovascular research, for example, has led the charge towards clinical translation of PRS [1]. This enthusiasm is driven by observations that a polygenic burden of LDL-increasing SNPs can confer monogenic-equivalent risk of cardiovascular disease, with polygenic scores improving clinical models for risk assessment and statin prescription that can reduce coronary heart disease and improve healthcare delivery efficiency [2,3,5]. However, many of these studies have been conducted exclusively in European descent populations, with few studies rigorously evaluating population-level applicability to non-Europeans. Those existing findings indeed demonstrate a large reduction in prediction utility in non-European populations [7], though often with comparisons of odds ratios

among arbitrary breakpoints in the risk distribution that make comparisons across studies challenging. To better clarify how polygenic prediction will be deployed in a clinical setting with diverse populations, more systematic and thorough evaluations of the utility of PRS within and across populations for many complex traits are still needed. These evaluations would benefit from rigorous polygenic prediction accuracy evaluations, especially for diverse non-European patients [50-52].

Translational genetic prediction may uniquely exacerbate disparities

Our impetus for raising these statistical issues limiting the generalizability of PRS across population stems from our concern that, while they are legitimately clinically promising for improving health outcomes for many biomedical phenotypes, they may have a larger potential to raise health disparities than other clinical factors for several reasons. The opportunities they provide for improving health outcomes means they inevitably will and should be pursued in the near term, but we urge that a concerted prioritization to make GWAS summary statistics easily accessible for diverse populations and a variety of traits and diseases is imperative, even when they are a fraction the size of the largest existing European datasets. *Individual* clinical tests, biomarkers, and prescription drug efficacy may vary across populations in their utility, but are fundamentally informed by the same underlying biology [53,54] Currently, guidelines state that as few as 120 individuals define reference intervals for clinical factors (though often smaller numbers from only one subpopulation are used) and there is no clear definition of who is “normal” [53]. Consequently, reference intervals for biomarkers can sometimes deviate considerably by reported ethnicity [55-57]. Defining ethnicity-specific reference intervals is

297 clearly an important problem that can provide fundamental interpretability gains with
298 implications for some major health benefits (e.g. need for dialysis and development of
299 Type 2 diabetes based on ethnicity-specific serum creatinine and hemoglobin A1C
300 reference intervals, respectively) [56]. Simply put, some biomarkers or clinical tests scale
301 directly with health outcomes independent of ancestry, and many others may have
302 distributional differences by ancestry but are equally valid after centering with respect to
303 a readily collected population reference.
304
305 In contrast, PRS are uniformly less useful in understudied populations due to
306 differences in genomic variation and population history [9,10]. No analogous solution of
307 defining ethnicity-specific reference intervals would ameliorate health disparities
308 implications for PRS or fundamentally aid interpretability in non-European populations.
309 Rather, as we and others demonstrate, PRS are unique in that even with multi-ethnic
310 population references, these scores are fundamentally less informative in populations
311 more diverged from GWAS study cohorts.
312
313 The clinical use and deployment of genetic risk scores needs to be informed by the
314 issues surrounding tests that currently would unequivocally provide much greater
315 benefit to the subset of the world's population which is already on the positive end of
316 healthcare disparities*. Conversely, African descent populations, which already endure
317 many of the largest health disparities across the globe, are often predicted marginally
318 better, if at all, compared to random (**Figure 4**). They are therefore least likely to benefit

* To maximally benefit all populations, the largest existing GWAS results should be used. Downsampling European GWAS for the sake of parity results in worse predictors for all individuals.

from improvements in precision healthcare delivery from genetic risk scores with existing data due to human population history and study biases. This is a major concern globally and especially in the U.S., which already leads other middle- and high-income countries in both real and perceived healthcare disparities [58]. Thus, we would strongly urge that any discourse on clinical use of polygenic scores include a careful, quantitative assessment of the economic and health disparities impacts on underrepresented populations that might be unintentionally introduced by the use of PRS and raise awareness about how to eliminate these disparities.

How do we even the ledger?

What can be done? An equal investment in GWAS across all major ancestries and global populations is the most obvious solution to truly generate a substrate for equally informative risk scores, but is not likely to occur any time soon absent a dramatic priority shift given the current imbalance and stalled diversifying progress over the last five years (**Figure 1**). While it may be challenging or in some cases infeasible to acquire sample sizes large enough for PRS to be equally informative in all populations, some much-needed efforts towards increasing diversity in genomics that support open sharing of GWAS summary data from multiple ancestries are underway. Examples include the *All of Us* Research Program, the Population Architecture using Genomics and Epidemiology (PAGE) Consortium, as well as some disease-focused consortia, such as the T2D-genes and Stanley Global initiatives on the genetics of type II diabetes and psychiatric disorders, respectively. The prerequisite data for dramatically increasing diversity also hypothetically exist in several large-scale publicly funded datasets such as

the Million Veterans Project and Trans-Omics for Precision Medicine (TOPMed), but with problematic data access issues in which even summary data from GWAS within and across populations are not publicly shared. While there is an understandable patient privacy balance to strike when sharing individual-level data, GWAS summary statistics by population from all publicly funded and as many privately funded projects as possible should be made easily and publicly accessible to improve global health outcomes. Efforts to unify phenotype definitions, normalization approaches, and GWAS methods among studies are also encouraged.

To enable progress towards parity, it will be critical that open data sharing standards be adopted for all ancestries and for genetic studies of all sample sizes, not just the largest European results. Locally appropriate and secure genetic data sharing techniques as well as equitable technology availability will need to be adopted widely in Asia and Africa as they are in Europe and North America, to ensure that maximum value is achieved from existing and ongoing efforts that are being developed to help counter the current imbalance. Methodological improvements that better define risk scores by accounting for population allele frequency, LD, and/or admixture differences appropriately are underway and may help considerably, but will not by themselves bring equality. All of these efforts are important and should be prioritized not just for risk prediction but more generally to maximize the use and applicability of genetics to inform on the biology of disease. Given the acute recent attention on clinical use of PRS, we believe it is paramount to recognize their potential to improve health outcomes for all individuals and many complex diseases. Simultaneously, we as a field must address the

365 disparity in utility in an ethically thoughtful and scientifically rigorous fashion, lest we
366 inadvertently enable genetic technologies to contribute to, rather than reduce, existing
367 health disparities.
368

369 **Acknowledgments**

370 We thank Amit Khera for helpful discussions. We also thank Michiaki Kubo, Yoshinori
371 Murakami, and Masato Akiyama for their support in the BioBank Japan Project analysis.
372 This work was supported by funding from the National Institutes of Health
373 (K99MH117229 to A.R.M.). UK Biobank analyses were conducted via application
374 31063. The BioBank Japan Project was supported by the Tailor-Made Medical
375 Treatment Program of the Ministry of Education, Culture, Sports, Science, and
376 Technology (MEXT) and the Japan Agency for Medical Research and Development
377 (AMED). M.K. was supported by a Nakajima Foundation Fellowship and the Masason
378 Foundation.
379

380 **Figures**

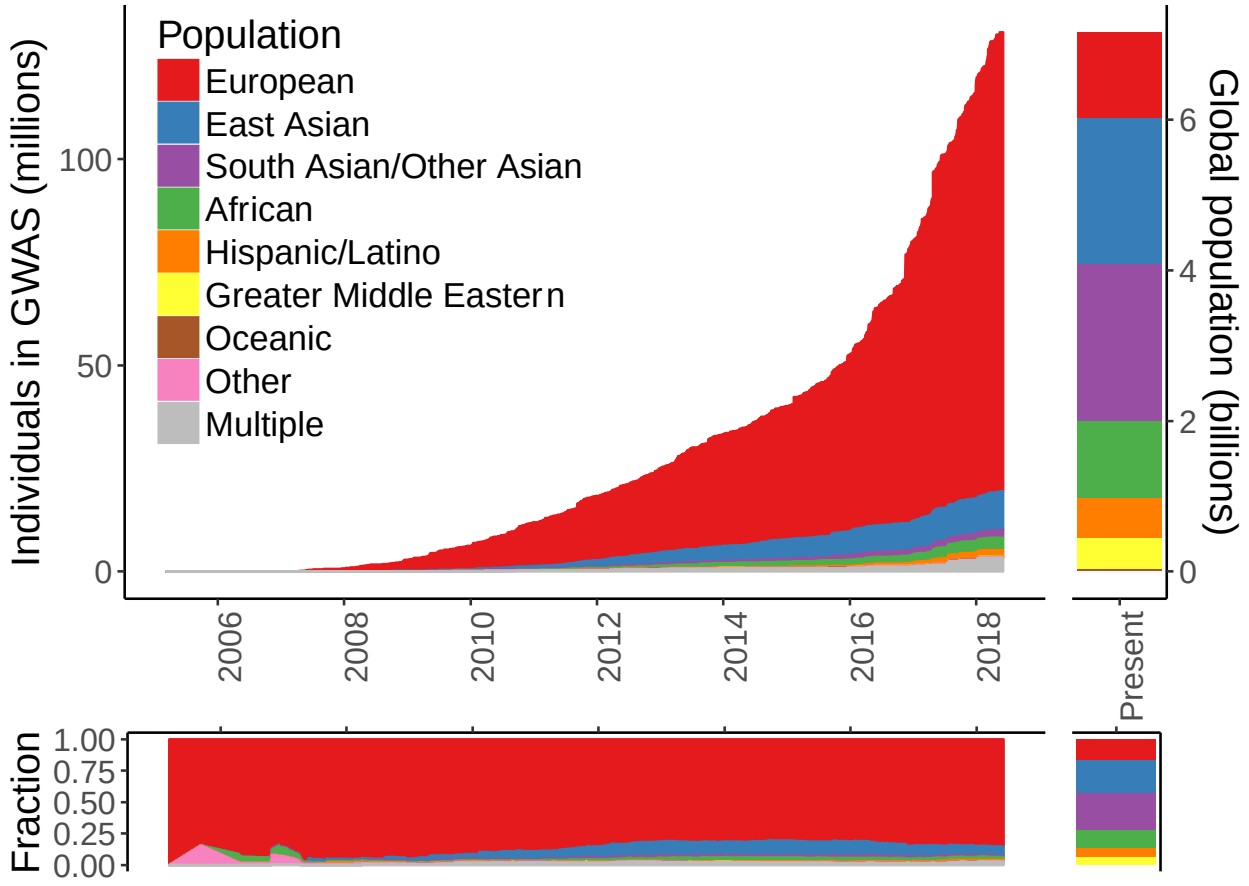
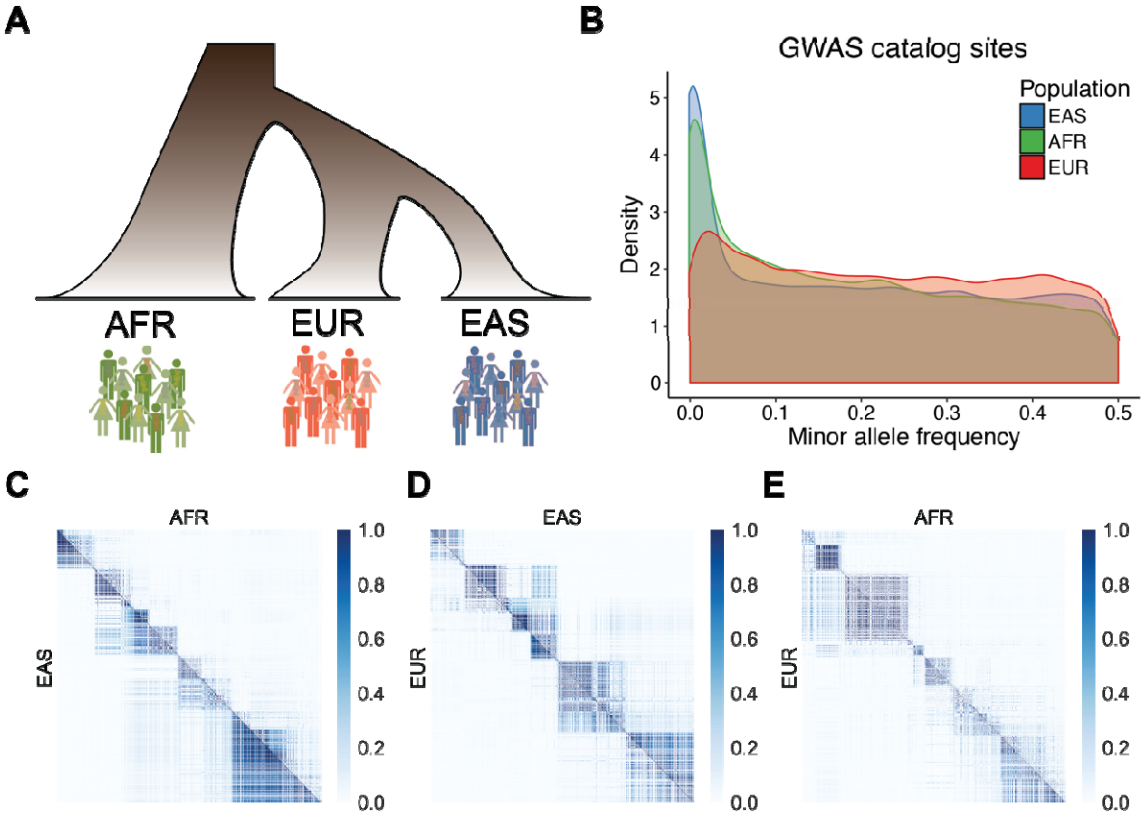


Figure 1 – Ancestry of GWAS participants over time compared to the global population. Cumulative data as reported by the GWAS catalog [23]. A notable caveat is that because some cohorts are included in numerous studies, some individuals are represented multiple times. This bias in multiple counting is especially likely for publicly available cohorts, which are more likely to be of European or East Asian descent. Individuals whose ancestry is “not reported” are not shown.



388

389 **Figure 2 – Demographic relationships, allele frequency differences, and local LD**

390 **patterns between population pairs.** Data analyzed from 1000 Genomes, in which

391 population labels are: AFR = continental African, EUR = European, and EAS = East

392 Asian. A) Cartoon relationships among AFR, EUR, and EAS populations. B) Allele

393 frequency distributions in AFR, EUR, and EAS populations of variants from the GWAS

394 catalog. C-E) Color axis shows LD scale (r^2). LD comparisons between pairs of

395 populations show the same region of the genome for each comparison (representative

396 region is chr1, 51572kb-52857kb) among pairs of SNPs polymorphic in both

397 populations, illustrating that different SNPs are polymorphic across some population

398 pairs, and that these SNPs have variable LD patterns across populations.

399

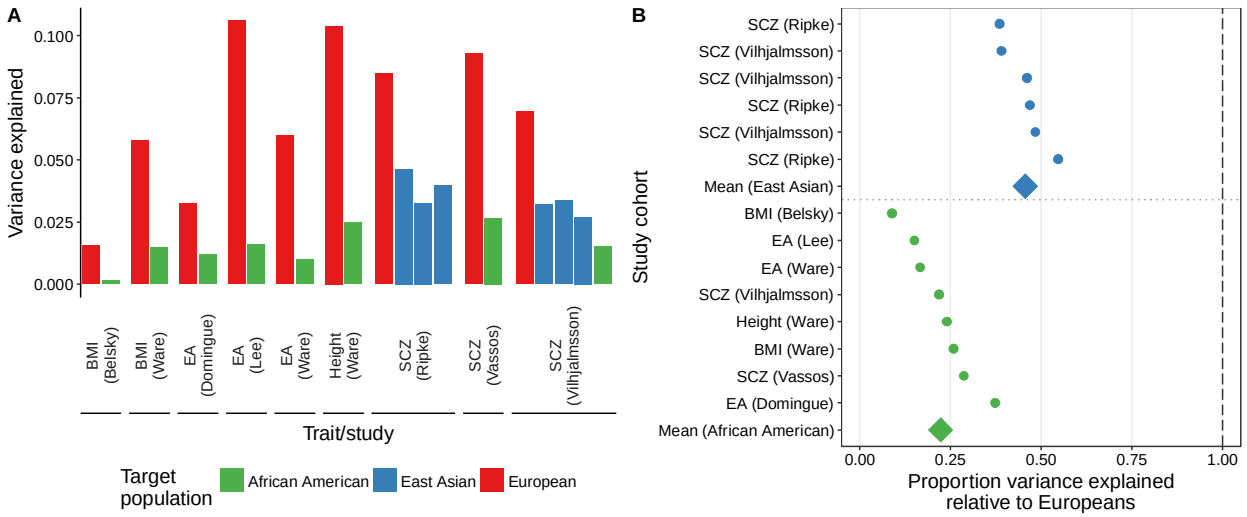


Figure 3 – Empirical comparison of phenotypic variance explained across populations using polygenic scores computed with European GWAS. All GWAS studies included here were conducted in European ancestry populations, with PRS calculated and evaluated in independent European, East Asian, and African American target cohorts. The European study biases result in the highest prediction accuracies in independent European cohorts, followed by declining accuracy with increased genetic divergence from Europe. A) Proportion of variance explained in each of the original studies. B) Relative proportion of variance explained in each population with respect to an independent European target population in each study. The diminished proportion of variance explained in East Asian and African American populations relative to Europeans is remarkably consistent despite differing genetic architectures, prediction methods, and accuracy metrics due to similar population histories within these cohorts. BMI = body mass index, EA = educational attainment, and SCZ = schizophrenia.

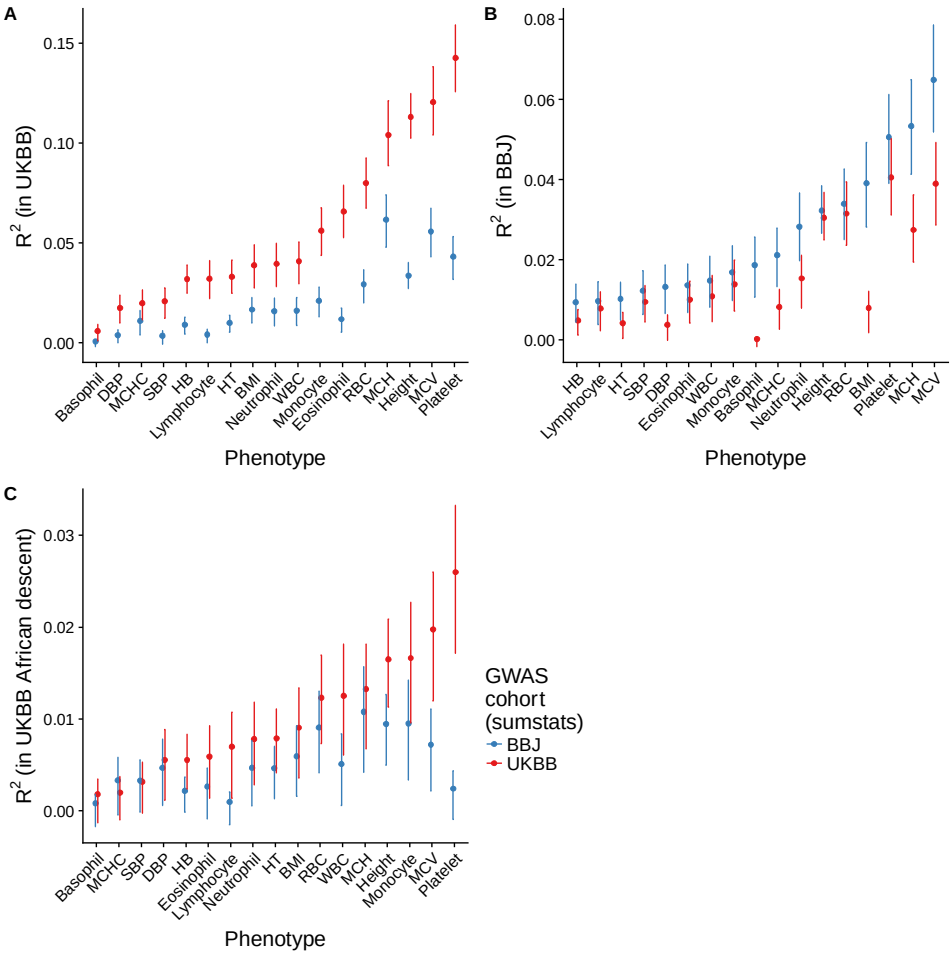


Figure 4 – Polygenic risk prediction accuracy in Japanese, British, and African descent individuals using independent GWAS of equal sample sizes in the BioBank Japan and UK Biobank. All target prediction cohorts are withheld from the GWAS and thus independent. Sample sizes in each GWAS are identical between BBJ and UKBB (**Table S1**). To optimize signal to noise, each point shows the maximum R^2 (i.e. best predictor) across 10 p-value thresholds. R^2 values for all p-value thresholds are shown in **Figures S2-S4**. Prediction accuracy tends to be higher in the UK Biobank, likely because observed heritability tends to be higher than in the BioBank Japan (**Table S3**). A) Genetic prediction accuracy for 17 anthropometric and blood panel traits in

Japanese individuals using summary statistics from GWAS of independent BioBank Japan versus UK Biobank samples. B) Genetic prediction accuracy for the same 17 anthropometric and blood panel traits in independent British individuals using summary statistics from GWAS of independent BioBank Japan versus the UK Biobank samples. C) Genetic prediction accuracy for 17 anthropometric and blood panel traits in African descent individuals in the UK Biobank using summary statistics from GWAS of independent BioBank Japan versus UK Biobank samples.

References

1. Knowles, J. W. & Ashley, E. A. Cardiovascular disease: The rise of the genetic risk score. *PLoS Med* **15,** e1002546–7 (2018).
2. Khera, A. V. *et al.* Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. (2017). doi:10.1101/218388
3. 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).
4. 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).
5. 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 SettingClinical Perspective. *Circulation* **135,** 2091–2101 (2017).
6. 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).
7. Khera, A. V. *et al.* Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med* **375,** 2349–2358 (2016).
8. Khera, A. V. & Kathiresan, S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet* **18,** 331–344 (2017).
9. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100,** 635–649 (2017).
10. Scutari, M., Mackay, I. & Balding, D. Using Genetic Distance to Infer the Accuracy of Genomic Prediction. *PLoS Genet* **12,** e1006288 (2016).
11. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics* **97,** 576–592 (2015).
12. Ware, E. B. *et al.* Heterogeneity in polygenic scores for common human traits. 1–13 (2017). doi:10.1101/106062
13. Curtis, D. Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *bioRxiv* (2018). doi:https://doi.org/10.1101/287136
14. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
15. Belsky, D. W. *et al.* Development and Evaluation of a Genetic Risk Score for Obesity. *Biodemography and Social Biology* **59,** 85–100 (2013).
16. Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M. & Boardman, J. D. Polygenic Influence on Educational Attainment. *AERA Open* **1,** 233285841559997–13 (2015).
17. Wedow, R. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* **92,** 109 (2018).

18. Vassos, E. *et al.* An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. *Biological Psychiatry* **81,** 470–477 (2017).
19. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat Genet* **49,** 1458–1467 (2017).
20. Li, Z. *et al.* Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nature Publishing Group* **49,** 1576–1583 (2017).
21. Need, A. C. & Goldstein, D. B. Next generation disparities in human genomics: concerns and remedies. *Trends in Genetics* **25,** 489–494 (2009).
22. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538,** 161–164 (2016).
23. Morales, J. *et al.* A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. 1–10 (2018). doi:10.1186/s13059-018-1396-2
24. Rosenberg, N. A. *et al.* Genome-wide association studies in diverse populations. *Nat Rev Genet* **11,** 356–366 (2010).
25. Sham, P. C., Cherny, S. S., Purcell, S. & Hewitt, J. K. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *The American Journal of Human Genetics* **66,** 1616–1630 (2000).
26. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
27. Consortium, T. S. 2. D. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* **506,** 97–101 (2014).
28. Estrada, K. *et al.* Association of a Low-Frequency Variant in HNF1A With Type 2 Diabetes in a Latino Population. *JAMA* **311,** 2305–2314 (2014).
29. Haiman, C. A. *et al.* Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nat Genet* **43,** 570–573 (2011).
30. Genovese, G. *et al.* Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* **329,** 841–845 (2010).
31. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47,** 979–986 (2015).
32. 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 (2013).
33. Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447,** 1087–1093 (2007).
34. Mahajan, A. *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).
35. 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 (2010).
36. Huang, H. *et al.* Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547,** 173–178 (2017).

37. Berg, J. J. *et al.* Reduced signal for polygenic adaptation of height in UK Biobank. 1–44 (2018). doi:10.1101/354951
38. 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
39. Duncan, L. *et al.* Analysis of Polygenic Score Usage and Performance across Diverse Human Populations. (2018). doi:10.1101/398396
40. Novembre, J. & Barton, N. H. Tread Lightly Interpreting Polygenic Tests of Selection. *Genetics* **208,** 1351–1355 (2018).
41. Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nat Rev Genet* **16,** 333–343 (2015).
42. 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).
43. Paré, G., Mao, S. & Deng, W. Q. A machine-learning heuristic to improve gene score prediction of polygenic traits. *Sci Rep* **7,** 12665 (2017).
44. Martin, A. R. *et al.* An Unexpectedly Complex Architecture for Skin Pigmentation in Africans. (2017). doi:10.1101/200139
45. Hindorff, L. A. *et al.* Prioritizing diversity in human genomics research. *Nature Publishing Group* 1–11 (2017). doi:10.1038/nrg.2017.89
46. H3Africa Consortium *et al.* Research capacity. Enabling the genomic revolution in Africa. *Science* **344,** 1346–1348 (2014).
47. Duncan, L. E. *et al.* Largest GWAS of PTSD (N=20□070) yields genetic overlap with schizophrenia and sex differences in heritability. *Molecular Psychiatry* (2017). doi:10.1038/mp.2017.77
48. Howrigan, D. Details and Considerations of the Uk Biobank GWAS. *http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas* (2017). Available at: (Accessed: 9 November 2017)
49. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* 1–16 (2018). doi:10.1038/s41588-018-0047-6
50. Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nature Publishing Group* **14,** 507–515 (2013).
51. Wray, N. R. *et al.* Research Review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatr* **55,** 1068–1087 (2014).
52. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 1–10 (2018). doi:10.1038/s41576-018-0018-x
53. Manrai, A. K., Patel, C. J. & Ioannidis, J. P. A. In the Era of Precision Medicine and Big Data, Who Is Normal? *JAMA* **319,** 1981–1982 (2018).
54. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nature Publishing Group* **12,** 581–594 (2013).
55. Carroll, M. D., Kit, B. K., Lacher, D. A., Shero, S. T. & Mussolino, M. E. Trends in lipids and lipoproteins in US adults, 1988-2010. *JAMA* **308,** 1545–1554 (2012).

569 56. Rappoport, N. *et al.* Creating ethnicity-specific reference intervals for lab tests
570 from EHR data. *bioRxiv* (2017). doi:10.1101/213892
571 57. Lim, E., Miyamura, J. & Chen, J. J. Racial/Ethnic-Specific Reference Intervals for
572 Common Laboratory Tests: A Comparison among Asians, Blacks, Hispanics, and
573 White. *Hawaii J Med Public Health* **74,** 302–310 (2015).
574 58. Hero, J. O., Zaslavsky, A. M. & Blendon, R. J. The United States Leads Other
575 Nations In Differences By Income In Perceptions Of Health And Health Care.
576 *Health Aff* **36,** 1032–1040 (2017).
577