Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations

Alicia R. Martin,1,2,5,4 Christopher R. Gignoux,4 Raymond K. Walters,1,2,3 Genevieve L. Wojcik,4 Benjamin M. Neale,1,2,3 Simon Gravel,5,6 Mark J. Daly,1,2,3 Carlos D. Bustamante,4 and Eimear E. Kenny7,8,9,10,*

The vast majority of genome-wide association studies (GWASs) are performed in Europeans, and their transferability to other populations is dependent on many factors (e.g., linkage disequilibrium, allele frequencies, genetic architecture). As medical genomics studies become increasingly large and diverse, gaining insights into population history and consequently the transferability of disease risk measurement is critical. Here, we disentangle recent population history in the widely used 1000 Genomes Project reference panel, with an emphasis on populations underrepresented in medical studies. To examine the transferability of single-ancestry GWASs, we published summary statistics to calculate polygenic risk scores for eight well-studied phenotypes. We identify directional inconsistencies in all scores; for example, height is predicted to decrease with genetic distance from Europeans, despite robust anthropological evidence that West Africans are as tall as Europeans on average. To gain deeper quantitative insights into GWAS transferability, we developed a complex trait coalescent-based simulation framework considering effects of polygenicity, causal allele frequency divergence, and heritability. As expected, correlations between true and inferred risk are typically highest in the population from which summary statistics were derived. We demonstrate that scores inferred from European GWASs are biased by genetic drift in other populations even when choosing the same causal variants and that biases in any direction are possible and unpredictable. This work cautions that summarizing findings from large-scale GWASs may have limited portability to other populations using standard approaches and highlights the need for generalized risk prediction methods and the inclusion of more diverse individuals in medical genomics.

Introduction

The majority of genome-wide association studies (GWASs) have been performed in populations of European descent.1–4 An open question in medical genomics is the degree to which these results transfer to new populations. GWASs have yielded tens of thousands of common genetic variants significantly associated with human medical and evolutionary phenotypes, most of which have replicated in other ethnic groups.5–7 However, GWASs are optimally powered to discover common variant associations, and the European bias in GWASs results in associated SNPs with higher minor allele frequencies on average compared to other populations. The predictive power of GWAS findings and genetic diagnostic accuracy in non-Europeans are therefore limited by population differences in allele frequencies and linkage disequilibrium structure. For example, a previous study showed that the accuracy of breeding values and genomic prediction decays approximately linearly with increasing divergence between the discovery and target population.8 Additionally, multiple individuals with African ancestry have received false positive misdiagnoses of hypertrophic cardiomyopathy that would have been prevented with the inclusion of even small numbers of African Americans in these studies.9 Further, a previous study finding that 96% of GWAS participants are of European descent7 has recently been updated; although the non-European proportion of GWAS participants has increased to nearly 20%, this is primarily driven by Asian individuals, and the proportion of individuals with African and Hispanic/Latino ancestry in GWASs has remained essentially unchanged.4 As GWAS sample sizes grow to hundreds of thousands of samples, they also become better powered to detect rare variant associations.10–12 Large-scale sequencing studies have demonstrated that rare variants show stronger geographic clustering than common variants.13–15 Rare, disease-associated variants are therefore expected to track with recent population demography and/or be population restricted.14,16–18 As the next era of GWASs expands to evaluate the disease-associated role of rare variants, it is not only scientifically imperative to include multi-ethnic populations, it is also likely that such studies will encounter increasing genetic heterogeneity in very large study populations. A comprehensive understanding of the genetic diversity and demographic history of multi-ethnic populations is critical for appropriate applications of GWASs and ultimately for ensuring that genetics does not contribute to or enhance health disparities.4

1Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA; 2Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; 3Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; 4Department of Genetics, Stanford University, Stanford, CA 94305, USA; 5Department of Human Genetics, McGill University, Montreal, QC H3A 0G1, Canada; 6McGill University and Genome Quebec Innovation Centre, Montreal, QC H3A 0G1, Canada; 7Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 8The Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 9Center of Statistical Genetics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 10Imb Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
*Correspondence: eimear.kenny@mssm.edu
http://dx.doi.org/10.1016/j.ajhg.2017.03.004.
© 2017 American Society of Human Genetics.

The American Journal of Human Genetics 100, 635–649, April 6, 2017

635
The most recent release of the 1000 Genomes Project (phase 3) provides one of the largest global reference panels of whole-genome sequencing data, enabling a broad survey of human genetic variation.\textsuperscript{19} The depth and breadth of diversity queried facilitates a deep understanding of the evolutionary forces (e.g., selection and drift) shaping existing genetic variation in present-day populations that contribute to adaptation and disease.\textsuperscript{20–25} Studies of admixed populations have been particularly fruitful in identifying genetic adaptations and risk for diseases that are stratified across diverged ancestral origins.\textsuperscript{26–31} Admixture patterns became especially complex during the peopling of the Americas, with extensive recent admixture spanning multiple continents. Processes shaping structure in these admixed populations include sex-biased migration and admixture, isolation-by-distance, differential drift in mainland versus island populations, and variable admixture timing.\textsuperscript{14,32,33}

Standard GWAS strategies approach population structure as a nuisance factor. A typical stepwise procedure first detects dimensions of global population structure in each individual, using principal-component analysis (PCA) or other methods,\textsuperscript{34–37} and often excludes “outlier” individuals from the analysis and/or corrects for inflation arising from population structure in the statistical model for association. Such strategies reduce false positives in test statistics, but can also reduce power for association in heterogeneous populations and are less likely to work for rare variant association.\textsuperscript{38,39} Recent methodological advances have leveraged patterns of global and local ancestry for improved association power,\textsuperscript{27,40,41} fine-mapping\textsuperscript{42} and genome assembly.\textsuperscript{43} At the same time, population genetic studies have demonstrated the presence of fine-scale sub-continental structure in the African, Native American, and European components of populations from the Americas.\textsuperscript{44–47} If trait-associated variants follow the same patterns of demography, then we expect that modeling sub-continental ancestry may enable their improved detection in admixed populations.

The dawn of the GWAS era saw limited success in identifying genome-wide significant loci associated with disease, and a major endeavor to better understand the genetic architecture of complex traits emerged. The peaks that met genome-wide significance typically did not explain a significant fraction of the phenotypic variance, and a major goal to estimate how many more signals remained yet to be discovered arose; this objective ushered in a wave of methodological development in heritability, linear mixed models, and polygenic risk prediction, as discussed and reviewed extensively elsewhere.\textsuperscript{11,48–56} Numerous complex traits have been studied with cohort sizes in the hundreds of thousands, and yet in each case there are many more signals that improve prediction accuracy than meet genome-wide significance.\textsuperscript{48,57–59} For example, including only genome-wide significant loci in the prediction of schizophrenia explains <3% of the phenotypic variance, whereas loci meeting the significance threshold that optimally balances signal versus noise (in this case, p ≤ 0.1) in the meta-analysis explains considerably more (>18%) of the phenotypic variance.\textsuperscript{11} Because the prediction accuracy, which is usually measured via prediction $R^2$, Nagelkerke’s $R^2$, or receiver operator curve AUC, of polygenic risk scores is currently low for most traits,\textsuperscript{56} genetic risk prediction is not clinically viable at present, but polygenic risk scores have nonetheless repeatedly proven valuable in research contexts across a multitude of complex traits\textsuperscript{11,48,60–65} and will become increasingly useful as GWAS sample sizes grow.\textsuperscript{59} Additionally, several methodological advancements to the standard approach have recently been undertaken.\textsuperscript{58,66–68}

In this study, we explore the impact of population diversity on the landscape of variation underlying human traits. We infer demographic history for the global populations in the 1000 Genomes Project, focusing particularly on admixed populations from the Americas, which are under-represented in medical genetic studies.\textsuperscript{4} We disentangle local ancestry to infer the ancestral origins of these populations. We link this work to ongoing efforts to improve study design and disease variant discovery by quantifying biases in clinical databases and GWASs in diverse and admixed populations. These biases have a striking impact on genetic risk prediction; for example, a previous study calculated polygenic risk scores for schizophrenia in East Asians and Africans based on GWAS summary statistics derived from a European cohort and found that prediction accuracy was reduced by more than 50% in non-European populations.\textsuperscript{62} To disentangle the role of demography on polygenic risk prediction derived from single-ancestry GWASs, we designed a coalescent-based simulation framework reflecting modern human population history and show that polygenic risk scores derived from European GWASs are biased when applied to diverged populations. Specifically, we identify reduced variance in risk prediction with increasing divergence from Europe reflecting decreased overall variance explained, and demonstrate that an enrichment of low-frequency risk and high-frequency protective alleles contribute to an overall protective shift in European inferred risk on average across traits. Our results highlight the need for the inclusion of more diverse populations in GWASs as well as genetic risk prediction methods improving transferability across populations.

Material and Methods

Ancestry Deconvolution

We used the phased haplotypes from the 1000 Genomes consortium. We phased reference haplotypes from 43 Native American samples from Mao et al.\textsuperscript{69} inferred to have >0.99 Native ancestry in ADMIXTURE using SHAPEIT2 (v.2.r778).\textsuperscript{70} then merged the haplotypes using scripts made publicly available. These combined phased haplotypes were used as input to the PopPhased version of RFMix v.1.5.4\textsuperscript{71} with the following flags: -w 0.2, -e 1, -n 5, --use-reference-panels-in-EM, --forward-backward EM. The node size of 5 was selected to reduce bias in
random forests resulting from unbalanced reference panel sizes (AFR panel N = 504, EUR panel N = 503, and NAT panel N = 43). We used the default minimum window size of 0.2 cM to enable model comparisons with previously inferred models using Tracts. We used 1 EM iteration to improve the local ancestry calls without substantially increasing computational complexity. We used the reference panel in the EM to take better advantage of the Native American ancestry tracts from the Hispanic/Latinos in the EM given the small NAT reference panel. We set the LWK, MSL, GWG, YRI, and ESN as reference African populations, the CEU, GBR, FIN, IBS, and TSI as reference European populations, and the samples from Mao et al. with inferred >0.99 Native ancestry as reference Native American populations, as in Abecasis et al.

Ancestry-Specific PCA
We performed ancestry-specific PCA, as described in Moreno-Estrada et al. The resulting matrix is not necessarily orthogonalized, so we subsequently performed singular value decomposition in python 2.7 using numpy. There were a number of minor outliers, as seen previously. There was one outlier (ASW individual NA20314) when analyzing the African tracts, which was expected as this individual has no African ancestry. There were eight outliers (PUR HG00731, PUR HG00732, ACB HG01880, ACB HG01882, PEL HG01944, ACB HG02497, ASW NA20320, ASW NA20321) when analyzing the European tracts. Some of these individuals had minimal European ancestry, had South or East Asian ancestry misclassified as European ancestry resulting from a limited 3-way ancestry reference panel, or were unexpected outliers. As described in the PCAmask manual, a handful of major outliers sometimes occur. AS-PCA is an iterative procedure, we therefore removed the major outliers for each sub-continental analysis and orthogonalized the matrix on this subset.

Tracts
The RFMix output was collapsed into haploid bed files, and “UNK” or unknown ancestry was assigned where the posterior probability of a given ancestry was <0.90. These collapsed haploid tracts were used to infer admixture timings, quantities, and proportions for the ACB and PEL (new to phase 3) using Tracts. Because the ACB have a very small proportion of Native American ancestry, we fit three 2-way models of admixture, including one model of single- and two models of double-pulse admixture events, using Tracts. In both of the double-pulse admixture models, the model includes an early mixture of African and European ancestry followed by another later pulse of either European or African ancestry. We randomized starting parameters and fit each model 100 times and compared the log-likelihoods of the model fits. The single-pulse and double-pulse model with a second wave of African admixture provided the best fits and reached similar log-likelihoods, with the latter showing a slight improvement in fit.

We next assessed the fit of nine different models in Tracts for the PEL, including several two-pulse and three-pulse models. Ordering the populations as NAT, EUR, and AFR, we tested the following models: ppp_ppp, ppp_ppx, ppx_xpx, ppx_xpp, ppx_ppx, ppx_xpp, ppx_xpx, ppx_xpp, and ppx_xpx_xpx, where the order of each letter corresponds with the order of populations given above, an underscore indicates a distinct migration event with the first event corresponding with the most generations before present, p corresponds with a pulse of the ordered ancestries, and x corresponds with no input from the ordered ancestries. We tested all nine models preliminarily three times, and for all models that converged and were within the top three models, we subsequently fit each model with 100 starting parameter randomizations.

Imputation Accuracy
Imputation accuracy was calculated using a leave-one-out internal validation approach. Two array designs were compared for this analysis: Illumina OmniExpress and Affymetrix Axiom World Array LAT. Sites from these array designs were subset from chromosome 9 of the 1000 Genomes Project Phase 3 release for admixed populations. After fixing these sites, each individual was imputed using the rest of the dataset as a reference panel.

Overall imputation accuracy was binned by minor allele frequency (0.5%–1%, 1%–2%, 2%–3%, 3%–4%, 4%–5%, 5%–10%, 10%–20%, 20%–30%, 30%–40%, 40%–50%) comparing the genotyped true alleles to the imputed dosages. A second round of analyses stratified the imputation by local ancestry diplotype, which was estimated as described earlier. Within each ancestral diplotype (AFR_AFR, AFR_NAT, AFR_EUR, EUR_EUR, EUR_NAT, EUR_NAT, NAT_NAT), imputation accuracy was again estimated within MAF bins.

Empirical Polygenic Risk Score Inferences
In the most standard approach, genetic risk scores for a target cohort are generated using genome-wide summary statistics from a discovery GWAS with a set of SNPs common to both studies. From this starting set of SNPs, a further reduced set of pruned, approximately independent SNPs are then identified through a greedy clumping algorithm. Typically, progressively larger sets of SNPs defined by a range of p value thresholds (e.g., p < 5 × 10^{-8}, 1 × 10^{-5}, 1 × 10^{-4}, 1 × 10^{-3}, 0.01, etc.) are evaluated to identify the best model balancing the signal to noise ratio to maximize phenotypic variance explained. Once the optimal significance threshold and the final set of pruned, approximately independent set of SNPs have been selected, a polygenic risk score for each individual in a target sample is computed as the sum of the count of risk alleles weighted by the effect size (e.g., log odds ratio).

To compute polygenic risk scores in the 1000 Genomes samples using summary statistics from previous GWASs, we first filtered to biallelic SNPs and removed ambiguous AT/GC SNPs from the integrated 1000 Genome call set. To get relatively independent associations when multiple significant p value associations are in the same region in a GWAS (i.e., in LD), we performed clumping in plink using the --clump flag for all variants with MAF < 0.01, which uses a greedy algorithm ordering SNPs by p value, then selectively removes SNPs within close proximity and LD in ascending p value order (i.e., starting with the most significant SNP). As a population cohort with similar LD patterns to the study sets, we used European 1000 Genomes samples (CEU, GBR, FIN, IBS, and TSI). To compute the polygenic risk scores, we considered all SNPs with p values ≤ 1 × 10^{-2} in the GWAS, a window size of 250 kb, and an R^2 threshold of 0.5 in Europeans to group SNPs. After obtaining the most significant, approximately independent signals (Table S4), we computed polygenic scores using the --score flag in plink.

Polygenic Risk Score Simulations
We simulated genotypes in a coalescent framework with msprime v.1.3 for chromosome 20 incorporating a recombination map of GRCh37 and an assumed mutation rate of 2 × 10^{-8} mutations / (base pair * generation). We used a demographic model previously...
Figure 1. Sub-continental Diversity and Origins of African, European, and Native American Components of Recently Admixed American Populations

(A) ADMIXTURE analysis at K = 3 focusing on admixed Americas samples, with the NAT, CEU, and YRI as reference populations.

(legend continued on next page)
We then define the total liability as

\[ L = \sqrt{h^2} \cdot Z_X + \sqrt{1 - h^2} \cdot Z_E \]

where the normal distribution is specified by the mean and standard deviation (as in python’s numpy package). For all other non-causal sites, the effect size is zero. We then define \( X \) as

\[ X = \sum_{i=1}^{m} g_i \beta_i \]

where \( g_i \) are the genotype states (i.e., 0, 1, or 2). To handle varying allele frequencies and potential weak LD between causal sites, to ensure a neutral model with random true polygenic risks with respect to allele frequencies, and to obtain the total desired variance, we normalize \( X \) as

\[ Z_X = \frac{X - \mu_X}{\sigma_X} \]

We then compute the true polygenic risk score as

\[ G = \sqrt{h^2} \cdot Z_X \]

such that the total variance of the scores is \( h^2 \). We also simulated environmental noise and standardize to ensure equal variance between normalized genetic and environmental effects before, defining the environmental effect \( E \) as

\[ E = N(0, 1 - h^2) \]

\[ Z_E = \frac{e - \mu_E}{\sigma_E} \]

\[ E = \sqrt{1 - h^2} \cdot Z_E \]

such that the total variance of the environmental effect is \( 1 - h^2 \). We then define the total liability as

\[ L = \sqrt{h^2} \cdot Z_X + \sqrt{1 - h^2} \cdot Z_E \]

\[ G = E + L \]

We assigned 10,000 European individuals at the most extreme end of the liability threshold “case” status assuming a prevalence of 5%. We randomly assigned 10,000 different European individuals “control” status. We ran a GWAS with these 10,000 European case subjects and 10,000 European control subjects, computing Fisher’s exact test for all sites with MAF > 0.01. As before for empirical polygenic risk score calculations from real GWAS summary statistics, we clumped these SNPs into LD blocks for all sites with \( p \leq 1 \times 10^{-2} \), and \( R^2 \leq 0.5 \) in Europeans within a window size of 250 kb. We used these SNPs to compute inferred polygenic risk scores as before, summing the product of the log odds ratio and genotype for the true polygenic risk in a cohort of 10,000 simulated European, African, and East Asian individuals (all not included in the simulated GWAS). We compared the true versus inferred polygenic risk scores for these individuals across varying complexities \((m = 200, 500, 1,000) \) and heritabilities \((h^2 = 0.33, 0.50, 0.67)\).

**Results**

**Genetic Diversity within and between Populations in the Americas**

We first assessed the overall diversity at the global and sub-continental level of the 1000 Genomes Project (phase 3) populations using ADMIXTURE and PCA (Figures S1 and S2). The six populations from the Americas demonstrate considerable continental admixture, with genetic ancestry primarily from Europe, Africa, and the Americas, recapitulating previously observed population structure. To quantify continental genetic diversity in these populations, we repeated the analysis using YRI, CEU, and NAT samples as reference panels (population labels and abbreviations in Table S1). We observed widely varying continental admixture contributions in the six populations from the Americas at \( K = 3 \) (Figure 1A and Table S2). For example, when compared to the ASW, the ACB have a higher proportion of African ancestry \((\mu = 0.88, 95\% CI = [0.87–0.89]) \) versus \( \mu = 0.76, 95\% CI = [0.73–0.78] \); two-sided t test \( p = 3.0 \times 10^{-13} \) and a smaller proportion of EUR and NAT ancestry. The PEL have more NAT ancestry than all of the other AMR populations \((\mu = 0.77, 95\% CI = [0.75–0.80]) \) versus CLM: \( \mu = 0.26, 95\% CI = [0.24, 0.27], p = 2.9 \times 10^{-95} \); PUR: \( \mu = 0.13, 95\% CI = [0.12, 0.13], p = 4.8 \times 10^{-93} \); and MXL: \( \mu = 0.47, 95\% CI = [0.43, 0.50], p = 1.7 \times 10^{-29} \) ascertained in 1000 Genomes.

We explored the origin of the subcontinental-level ancestry from recently admixed individuals by identifying local ancestry tracts (Material and Methods, Figure S3). As proxy sources of populations for the recent admixture, we used EUR and AFR continental samples from the 1000 Genomes Project as well as NAT samples genotyped previously. Concordance between global ancestry estimates inferred using ADMIXTURE at \( K = 5 \) and RFMix was typically high (Pearson’s correlation $\geq 98\%$, see Figure S4). Using Tracts, we modeled...
the length distribution of the AFR, EUR, and NAT tracts to infer that admixing began ~12 and ~8 generations ago in the PEL and ACB populations, respectively (Figure S5), consistent with previous estimates from other populations from the Americas.44,72,32

We further investigated the subcontinental ancestry of admixed populations from the Americas one ancestry at a time using a version of PCA modified to handle highly masked data (ancestry-specific or AS-PCA) as implemented in PCAmask.32 Example ancestry tracts in a PEL individual subset to AFR, EUR, and NAT components are shown in Figures 1B, 1D, and 1F, respectively. Consistent with previous observations, the inferred European tracts in Hispanic/Latino populations most closely resemble southern European IBS and TSI populations with some additional drift32 (Figure 1B). The European tracts of the PUR are more differentiated compared to the CLM, MXL, and PEL populations, consistent with sex bias (Figure S6 and Table S3) and excess drift from founder effects in this island population.32 In contrast to the southern European tracts from the Hispanic/Latino populations, the African descent populations in the Americas have European admixture that more closely resembles the northwestern CEU and GBR European populations. The clusters are less distinct, owing to lower overall fractions of European ancestry, but the European components of the Hispanic/Latino and African American populations are significantly different (Wilcoxon rank sum test \( p = 2.4 \times 10^{-69} \)).

The ability to localize aggregated ancestral genomic tracts enables insights into the evolutionary origins of admixed populations. To disentangle whether the considerable Native American ancestry in the ASW individuals arose from recent admixture with Hispanic/Latino individuals or recent admixture with indigenous Native American populations, we queried the European tracts. We find that the European tracts of all ASW individuals with considerable Native American ancestry are well within the ASW cluster and project closer in Euclidean distance with AS-PC1 and AS-PC2 to northwestern Europe than the European tracts from Hispanic/Latino samples (\( p = 1.15 \times 10^{-34} \)), providing support for the latter hypothesis and providing regional nuance to previous findings.44

We also investigated the African origin of the admixed AFR/AMR populations (ACB and ASW), as well as the Native American origin of the Hispanic/Latino populations (CLM, MXL, PEL, and PUR). The African tracts of ancestry from the AFR/AMR populations project closer to the YRI and ESN of Nigeria than the GWD, MSL, and LWK populations (Figure 1C). This is consistent with slave records and previous genome-wide analyses of African Americans indicating that most sharing occurred in West and West-Central Africa.80–82 There are subtle differences between the African origins of the ACB and ASW populations (e.g., difference in distance from YRI on AS-PC1 and AS-PC2 \( p = 6.4 \times 10^{-6} \)), likely due either to mild island founder effects in the ACB samples or differences in African source populations for enslaved Africans who remained in Barbados versus those who were brought to the USA. The Native tracts of ancestry from the AMR populations first separate the southernmost PEL populations from the CLM, MXL, and PUR on AS-PC1, then separate the northernmost MXL from the CLM and PUR on AS-PC2, consistent with a north-south cline of divergence among indigenous Native American ancestry (Figure 1G).32,83

**Impact of Continental and Sub-continental Diversity on Disease Variant Mapping**

To investigate the role of ancestry in phenotype interpretation from genetic data, we assessed diversity across populations and local ancestries for recently admixed populations across the whole genome and sites from two reference databases: the GWAS catalog and ClinVar pathogenic and likely pathogenic sites. We recapitulate results showing that there is less variation across the genome (both genome-wide and on the Affymetrix 6.0 GWAS array sites used in local ancestry calling) in out-of-Africa versus African populations, but that GWAS variants are more polymorphic in European and Hispanic/Latino populations (Figures S7A, S7B, S8A, and S8B). We use a normalized measure of the minor allele frequency, an indicator of the amount of diversity captured in a population, to obtain a background coverage of each population, as done previously (e.g., Figure S4 from Auton et al.19). We show that the Affymetrix 6.0 array has a slight European bias (Figures S5A and S6A). We compared the site frequency spectrum of variants across the genome versus at GWAS catalog sites and identify elevated allele frequencies at GWAS catalog loci, particularly in populations with more European ancestry (e.g., the EUR, AMR, and SAS super populations, Figures S5C and S5D). We further compared heterozygosity (estimated here as \( 2pq \)) and the site frequency spectrum in recently admixed populations across diploid and haploid local ancestry tracts, respectively. Sites in the GWAS catalog and ClinVar are more and less common than genome-wide variants, respectively (Figure 2). Whereas heterozygosity across the whole genome is highest in African ancestry tracts, it is consistently the greatest in European ancestry tracts across these databases (Figures 2, S8C, and S8D), reflecting a strong bias toward European study participants.1,4,19,84 These results highlight imbalances in genome interpretability across local ancestry tracts in recently admixed populations and the utility of analyzing these variants jointly with these ancestry tracts over genome-wide ancestry estimates alone.

We also assessed imputation accuracy across the 3-way admixed populations from the Americas (CLM, MXL, PUR) for two arrays: the Illumina OmniExpress and the Affymetrix Axiom World Array LAT. Imputation accuracy was estimated as the correlation \( (r^2) \) between the original genotypes and the imputed dosages. For both array designs, imputation accuracy across all minor allele frequency (MAF) bins was highest for populations with the largest proportion of European ancestry (PUR) and
Heterozygosity, estimated here as 2pq, is calculated in admixed populations stratified by diploid local ancestry in (A) the whole genome, (B) sites from the GWAS catalog, and (C) sites from ClinVar classified as “pathogenic” or “likely pathogenic.” The mean and 95% confidence intervals were calculated by bootstrapping 1,000 times. Populations not shown in a given panel have too few diploid ancestry tracts overlapping sites to calculate heterozygosity.

Transferability of GWAS Findings across Populations

To quantify the transferability of European-biased genetic studies to other populations, we next used published GWAS summary statistics to infer polygenic risk scores across populations for well-studied traits, including height, waist-hip ratio, schizophrenia, type II diabetes, and asthma (Figures 4A–4D and S11, Material and Methods). Most of these summary statistics are derived from studies with primarily European cohorts, although GWASs of type II diabetes have been performed in both European-specific cohorts as well as across multi-ethnic cohorts. We identify clear directional inconsistencies in these inferred scores. For example, although the height summary statistics show the expected southern/northern cline of increasing European height (FIN, CEU, and GBR populations have significantly higher polygenic risk scores than IBS and TSI, p = 1.5 x 10^-75, Figure S9A), polygenic scores for height across super populations show biased predictions; the African populations sampled are genetically predicted to be considerably shorter than all Europeans and minimally taller than East Asians (Figure 4A), which contradicts empirical observations (with the exception of some indigenous pygmy/pygmoid populations). Additionally, polygenic risk scores for schizophrenia, while at a similar prevalence across populations where it has been well studied and sharing significant genetic risk across populations, shows considerably decreased scores in Africans compared to all other populations (Figure 4B). Lastly, the relative order of polygenic risk scores computed for type II diabetes across populations differs depending on whether the summary statistics are derived from a European-specific (Figure 4C) or multi-ethnic (Figure 4D) cohort.

Ancestry-Specific Biases in Polygenic Risk Score Estimates

We performed coalescent simulations to determine how GWAS signals discovered in one ancestral case/control cohort (i.e., “single-ancestry” GWAS) are expected to impact polygenic risk score estimates in other populations under neutrality using summary statistics (for details, see Material and Methods). In brief, we simulated variants according to a previously published demographic model inferred from Africans, East Asians, and Europeans. We
specified “causal” alleles and effect sizes randomly, such that each causal variant has evolved neutrally and has a mean effect of zero with the standard deviation equal to the global heritability divided by number of causal variants. We computed the true polygenic risk for each individual as the product of the estimated effect sizes and genotypes, then standardized the scores across all individuals. We calculated the total liability as the sum of the genetic and random environmental contributions, then identified 10,000 European case subjects with the most extreme liabilities and 10,000 other European control subjects. We computed Fisher’s exact tests with this European case-control cohort, then quantified inferred polygenic risk scores as the sum of the product of genotypes and log odds ratios for 10,000 samples per population not included in the GWAS.

In our simulations and consistent with realistic coalescent models, most variants are rare and population specific; “causal” variants are sampled from the global site frequency spectrum, resulting in subtle differences in true polygenic risk across populations (Figures S12, 5A, and 5B). We mirrored standard practices for performing a GWAS and computing polygenic risk scores (see above and Material and Methods). While causal variants in our simulations are drawn from the global site frequency spectrum and are therefore mostly rare, inferred scores are derived specifically from common variants that are typically much more common in the study population than elsewhere (here Europeans with case/control MAF ≥ 0.01). Consequently, while the distribution of mean true polygenic risk across simulation runs for each population are not significantly different (Figure 5A), the inferred risk is less than zero in Europeans ($p = 1.9 \times 10^{-54}$, 95% CI $= [-84.3, -67.4]$), slightly less than zero in East Asians ($p = 5.9 \times 10^{-5}$, 95% CI $= [-19.1, -6.6]$), and not significantly different from zero in Africans (Figure 5B); the variance in inferred risk scores, a proxy for the fraction of heritable variation explained, also decreases with this trend. Specifically, when $h^2 = 0.67$ and $m = 1,000$ causal markers, we find that the true and inferred polygenic risk scores in the EUR population are significantly correlated (i.e., non-zero, mean $r = 0.59$, $p < 1 \times 10^{-200}$), but the correlations in EAS and AFR populations are significantly less than in EUR ($r = 0.35$ and $p = 1.5 \times 10^{-48}$, $r = 0.22$ and $p < 1 \times 10^{-200}$, respectively). Because of allele frequency differences, number of SNPs, and inferred effect size differences along the frequency spectrum, the scale is orders of magnitude different between the true and inferred raw, unstandardized scores, cautioning that while they are informative on a relative scale (Figures 5C and S11), their absolute scale should not be over interpreted. The inferred risk difference between populations is driven by the increased power to detect minor risk alleles rather than protective alleles in the study population,93 given the differential selection of case and control subjects in the liability threshold model. We demonstrate this empirically in these neutral simulations within the European population (Figure S14A), indicating that this phenomenon occurs even in the absence of population structure and when case and control cohort sizes are equal.
We find that the correlation between true and inferred polygenic risk is generally low (Figures 5C and S13), consistent with limited variance explained by polygenic risk scores from GWASs of these cohort sizes for height (e.g., ~10% of variance explained for a cohort of size 183,72763) and schizophrenia (e.g., ~7% variance explained for a cohort of size 36,989 case subjects and 113,075 control subjects11). Low correlations in our simulations are most likely because common tag variants are a poor proxy for rare causal variants. As expected, correlations between true and inferred risk within populations are typically highest in the European population (i.e., the population in which variants were discovered, Figures 5A and S13). To quantify the differential prediction accuracy of polygenic risk scores across populations, we also evaluate the log odds ratio of being a case subject compared to a control subject across deciles of inferred polygenic risk in each population. We identify greater power to discern between case and control subjects in the EUR discovery population relative to the AFR and EAS populations (i.e., more heritable variation explained, as evidenced by a steeper slope) (Figure S14B). Across all populations, the mean Spearman correlations between true and inferred polygenic risk increase with increasing heritability while the standard deviations of these correlations significantly decrease (p = 0.05); however, there is considerable within-population heterogeneity resulting in high variation in scores across all populations. We find that in these neutral simulations, a polygenic risk score bias in essentially any direction is possible even when choosing the exact same causal variants and heritability and varying only fixed effect size (i.e., inferred polygenic risk in Europeans can be higher, lower, or intermediate compared to true risk relative to East Asians or Africans, Figures S12 and 5B).

Discussion

To date, GWASs have been performed opportunistically in primarily single-ancestry European cohorts, and an open question remains about their biomedical relevance for disease associations in other ancestries. As studies gain power by increasing sample sizes, effect size estimates become more precise and novel associations at lower frequencies are feasible. However, rare variants are largely population-private, and their effects are unlikely to transfer to new populations. Because linkage disequilibrium and allele frequencies vary across ancestries, effect size estimates from diverse cohorts are typically more precise than from single-ancestry cohorts (and often tempered),5 and the resolution of causal variant fine-mapping is considerably improved.87 Across a range of genetic architectures, diverse cohorts provide the opportunity to reduce false positives. At the Mendelian end of the spectrum, for example, disentangling risk variants with incomplete penetrance from benign false positives and localizing functional effects in genes is much more feasible with large diverse population cohorts than with single-ancestry analyses.94 Multiple false positive reports of pathogenic variants causing hypertrophic cardiomyopathy, a disease with relatively simple genomic architecture, have been returned to individuals of African descent or unspecified ancestry that would have been prevented if even a small number of African American samples were included in control cohorts.9 At the highly complex end of the polygenicity spectrum, we and others have shown that the utility of polygenic risk inferences and...
the heritable phenotypic variance explained in diverse populations is improved with more diverse cohorts.92,95

Standard single-ancestry GWASs typically apply linear mixed model approaches and/or incorporate principal components as covariates to control for confounding from population structure with primarily European-descent cohorts.1–3 A key concern when including multiple diverse populations in a GWAS is that there is increasing likelihood of identifying false positive variants associated with disease that are driven by allele frequency differences across ancestries. However, previous studies have analyzed association data for diverse ancestries and replicated findings across ethnicities, assuaging these concerns.6,87 In this study, we show that this ancestry stratification is not continuous along the genome: long tracts of ancestrally diverse populations present in admixed samples from the Americas are easily and accurately detected.

Querying population substructure within these tracts recapitulates expected trends, e.g., European ancestry in African Americans primarily descends from northern Europeans in contrast to European ancestry from Hispanic/Latinos, which primarily descends from southern Europeans, as seen previously.44 Additionally, population substructure follows a north-south cline in the Native component of Hispanic/Latinos, and the African component of admixed African descent populations in the Americas most closely resembles reference populations from Nigeria (notwithstanding the limited set of African populations from the 1000 Genomes Project). Admixture mapping has been successful at large sample sizes for identifying ancestry-specific genetic risk factors for disease.30 Given the level of accuracy and subcontinental resolution attained with local ancestry tracts in admixed populations, we emphasize the utility

Figure 5. Coalescent Simulation Framework to Generate True and Inferred Polygenic Risk Scores

Results of true and inferred polygenic risk scores, as well as their correlation, were computed via GWAS summary statistics from 10,000 simulated EUR case and control subjects modeling European, East Asian, and African population history (demographic parameters are from Gravel et al.14).

(A) The distribution of mean true, unstandardized polygenic risk scores for each population across 500 simulations with \( m = 1,000 \) causal variants and \( h^2 = 0.67 \).

(B) The distribution of mean inferred, unstandardized polygenic risk for the same simulation parameters as in (A) (center) and standardized true versus inferred polygenic risk scores for three different coalescent simulation replicates showing 10,000 randomly drawn samples from each population not included as case or control subjects (right).

(C) Violin plots show Pearson’s correlation across 50 iterations per parameter set between true and inferred polygenic risk scores across differing genetic architectures, including \( m = 200, 500, \) and \( 1,000 \) causal variants and \( h^2 = 0.67 \).
of a unified framework to jointly analyze genetic associations with local ancestry simultaneously.40

The transferability of GWASs is aided by the inclusion of diverse populations.36,37 We have shown that European discovery biases in GWASs are recapitulated in local ancestry tracts in admixed samples. We have quantified GWAS study biases in ancestral populations and shown that GWAS variants are at lower frequency specifically within African and Native tracts and higher frequency in European tracts in admixed American populations. Imputation accuracy is also stratified across diverged ancestries, including across local ancestries in admixed populations. With decreased imputation accuracy especially on Native American tracts, there is decreased power for potential ancestry-specific associations. This differentially limits conclusions for GWASs in an admixed population in a two-pronged manner: the ability to capture variation and the power to estimate associations.

As GWASs scale to sample sizes on the order of hundreds of thousands to millions, genetic risk prediction accuracy at the individual level improves.39 However, we show that the utility of polygenic risk scores computed using GWAS summary statistics are dependent on genetic similarity to the discovery cohort. Best linear unbiased prediction (BLUP) methods have been proposed to improve risk scores, but they require access to raw genetic data typically from very large datasets, are also dependent on LD structure in the study population, and offer only modest improvements in prediction accuracy.52 Furthermore, polygenic risk scores (PRSs) contain a mix of true positives (which have the bias described above) and false positives in the training GWAS. False positives, being chance statistical fluctuations, do not have the same allele frequency bias and therefore unfortunately play an outsized role in applying a PRS in a new population.

We have demonstrated that polygenic risk scores computed via current standard methods with summary statistics from a single-ancestry discovery cohort have numerous problems: differences in polygenic risk scores across populations are significant but not supported by epidemiological or anthropometric studies of the same traits, and directionality biases in polygenic risk scores across populations are unpredictable. Our coalescent simulations recapitulate these results and show that across replicates (i.e., traits, and thus not necessarily within a single trait), cross-population prediction accuracy is diminished with increasing divergence from the discovery cohort. These simulations provide further insight into directional inconsistencies in inferred polygenic risk scores with the same demographic model across replicate simulations, indicating that different traits are likely to suffer from biases that cannot be adjusted, e.g., using principal components alone. Directional selection is expected to bias polygenic risk inferences even more. Because biases arise from genetic drift alone, we recommend (1) avoiding interpretations from polygenic risk score differences extrapolated across populations, as these are likely confounded by latent population structure that is not properly corrected for with current standard methods, (2) mean-centering polygenic risk scores for each population, and (3) computing polygenic risk scores in populations with similar demographic histories as the study sample to ensure maximal predictive power. Further, additional methods that account for local ancestry in genetic risk prediction to incorporate different ancestral linkage disequilibrium and allele frequencies are needed. This study demonstrates the utility of disentangling ancestry tracts in recently admixed populations for inferring recent demographic history and identifying ancestry-stratified analytical biases; we also motivate the need to include more ancestor-stratified cohorts in GWASs to ensure that health disparities arising from genetic risk prediction do not become pervasive in individuals of admixed and non-European descent.

Supplemental Data
Supplemental Data include 14 figures and 4 tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2017.03.004.

Conflicts of Interest

Acknowledgments
We thank Suyash Shringarpure, Brian Maples, Andres Moreno-Estrada, Danny Park, Noah Zaitlen, Alexander Gusev, and Alkes Price for helpful discussions/feedback. We thank Vemeri Antilla for providing GWAS summary statistics. We thank Jerome Kelleher for several conversations about msprime, providing example scripts, and implementing new simulation capabilities. This work was supported by funds from several grants: the National Human Genome Research Institute under award numbers U01HG009080 (E.E.K., C.D.B., C.R.G.), U01HG007419 (C.D.B., C.R.G., G.L.W.), U01HG007417 (E.E.K.), U01HG005208 (M.J.D.), T32HG000044 (C.R.G.), and R01GM083606 (C.D.B.), the National Institute of General Medical Sciences under award number T32GM007790 (A.R.M.) at the National Institute of Health, the National Institute for Mental Health U01MH094432-02 (R.G.W., M.J.D.), the Directorate of Mathematical and Physical Sciences award 1201234 (S.G., C.D.B.) at the National Science Foundation, the Canadian Institutes of Health Research through the Canada Research Chair program and operating grant MOP-136855 (S.G.), and a Sloan Research Fellowship (S.G.).

Received: November 23, 2016
Accepted: March 10, 2017
Published: March 30, 2017
References


Supplemental Data

Human Demographic History Impacts

Genetic Risk Prediction across Diverse Populations

Alicia R. Martin, Christopher R. Gignoux, Raymond K. Walters, Genevieve L. Wojcik, Benjamin M. Neale, Simon Gravel, Mark J. Daly, Carlos D. Bustamante, and Eimear E. Kenny
ADMIXTURE analysis at K=5, K=7, and K=8. K=8 has the lowest 10-fold cross-validation error of K=3-12. At K=5, this analysis separates continental ancestries in the super populations (AFR, AMR, EAS, EUR, and SAS, population abbreviations in Table S1). These results also highlight sub-continental substructure; for example, there is detectable substructure resembling European (EUR) and East Asian (EAS) ancestries in the SAS populations (population means range from 6.1-15.9% and 0.3-12.2%, respectively), with the highest rates of East Asian-like ancestry in the Bengalis from Bangladesh (BEB). In contrast, the greatest quantity of European-like ancestry in the SAS populations is in the Punjabi from Lahore, Pakistan (PJL), who are geographically the closest to Europe. Ancestral clines have been observed along geographical, caste, and linguistic axes in more densely sampled studies of South Asia.\textsuperscript{1,2} Increasing the model to K=6 there is also an east-west cline among African populations, while at K=7 we observe the north-south cline of European ancestry.\textsuperscript{3} While there is minimal Native American ancestry (<1%) in most African Americans across the United States, there is a substantial enrichment in several ASW individuals from 1000 Genomes (mean of 3.1%, and 9 samples with >5%, including NA19625, NA19921, NA20299, NA20300, NA20314, NA20316, NA20319, NA20414, and NA20274).\textsuperscript{4,5} Interestingly, one ASW individual has no African ancestry (NA20314, EUR= 0.40, NAT=0.59) but is the mother of NA20316 in an ASW duo with few Mendelian inconsistencies that suggest that the father mostly likely has ~80% African and ~20% European ancestry, similar to other ASW individuals. We also find evidence of East Asian admixture in several PEL samples (39% in HG01944, 12% in HG02345, 6% in HG0192, 5% in HG01933, and 5% in HG01948). Consistent with the autosomal evidence, the Y chromosome haplogroup for HG01944 (Q1a-M120) clusters most closely with two KHV samples and other East Asians rather than the Q-L54 subgroup expected in samples from South America.\textsuperscript{6}
Figure S2 – Principal components analysis of all samples showing the relative homogeneity of AFR, EUR, EAS, and SAS continental groups and continental mixture of admixed samples from the Americas (ACB, ASW, CLM, MXL, PEL, and PUR).
**Figure S3** – Schema of local ancestry calling pipeline
**Figure S4** – Concordance between global ancestry estimates across individuals via Pearson’s correlation from ADMIXTURE at K=5 as in Figure S1 versus 3-way RFMix inferences for AFR, EUR, and NAT ancestries. The correlation between ADMIXTURE and global ancestry estimates from RFMix was lower when there was minimal ancestry from a given source population and/or tracts were very short (<5 cM), e.g. NAT ancestry in the ACB (\(\rho=0.79\)) and AFR ancestry in the MXL (\(\rho=0.94\)). A) ACB. Substantial differences occurred in 1 ACB individual, HG01880, where considerable South Asian ancestry (31.8%) was classified as European ancestry due to limitations of the 3-way local ancestry reference panel. B) ASW. C) CLM. D) MXL. E) PEL. Substantial differences occurred in 2 PEL individuals, HG01944 and HG02345, where considerable East Asian ancestry (38.2% and 12.3%, respectively) was classified in RFMix as EUR and NAT ancestry due to limitations of the 3-way local ancestry reference panel. F) PUR.
Figure S5 – Demographic reconstruction through genetically dated recent admixture events in the Americas. A-B) Local ancestry tract length decay of AFR, EUR, and NAT continental ancestry tracts for the A) PEL and B) ACB. Points represent the observed distribution of ancestry tracts, and solid lines represent the distribution of the best-fit Markov model inferred using Tracts, with the shaded areas indicating one standard deviation confidence intervals. C-D) Admixture time estimates in number of generations ago, relative quantity of migrants, and ancestry proportions over time under the best-fitting model for the C) PEL and D) ACB. C) The best-fit model for the PEL begins ~12 generations ago, which is slightly more recent than for insular and Caribbean mainland populations. For example, admixture in Colombian and Honduran mainland populations was previously inferred to have begun 14 generations ago, whereas admixture in Cuban, Puerto Rican, Dominican, and Haitian populations began 16-17 generations ago. D) The best-fit model for the ACB was an initial pulse of admixture between Europeans and Africans followed by a later pulse of African ancestry. The best model indicates that admixture in the ACB began ~8 generations ago with the initial pulse containing 87.4% African ancestry and 12.6% European ancestry. The second pulse of African ancestry began ~5 generations ago and had only a minor overall contribution (4.4% of total pulse ancestry), which is consistent with either a later small pulse of African ancestry or movement of populations within the Caribbean. The admixture events we infer in the ACB are more recent than previous ASW and African American two-pulse models, which estimated that admixture began ~10-11 generations ago. Potential explanations for this small difference include differences in the ages of individual between the two cohorts and the fact that pulse timings indicate the generations that admixture most likely spanned rather than the exact generation during which admixture began.
Figure S6 – Comparison of ploidy-adjusted ADMIXTURE ancestry estimates obtained on the autosomes and X chromosome at K=3 with CEU, YRI, and NAT\(^9\) reference samples. 700,093 SNPs on the autosomes and 10,503 SNPs on the X chromosome were used to infer ancestry proportions. A) African descent and B) Hispanic/Latino samples. Sex-biased admixture has previously been shown to be ubiquitous in the Americas, impacting phenotypes strongly correlated with ancestry, such as pigmentation.\(^7,10-14\) We inferred sex-biases in admixture events by separately querying ploidy-adjusted admixture proportions on the X chromosome versus the autosomes, as previously described\(^10\). We computed 3-way admixture proportions for AMR and AFR/AMR via ADMIXTURE\(^15\) and consistently find across all six admixed AMR populations that the ratio of European ancestry is significantly depleted on the X chromosome compared to the autosomes, indicating a ubiquitous excess of breeding European males in the Americas, as seen previously\(^4,12,16\); there is also a significant excess of Native American ancestry (p<1e-2, Table S3) on the X chromosome in each of the AMR populations (p < 1e-4).
Figure S7 – Genetic variation and allele frequencies in global populations across all sites and at GWAS sites. A-B) GWAS study bias in European and American samples compared at all Affy6 sites from which local ancestry calls were made. All standardizations are computed as the ratio of minor alleles to total alleles per population minus the mean ratio across all individuals, then all divided by the standard deviation of this ratio. Error bars shows the standard error of the mean. A) Standardized across all Affy6 sites. B) Standardized across the intersection of Affy6 sites and the GWAS catalog. C-D) Allele frequencies within all super populations. Minor allele frequency fraction across C) all sites Affy6 sites, and D) the intersection of all Affy6 and GWAS catalog sites.
Figure S8 – Genetic variation in global and admixed populations across all sites and at GWAS sites. A-B) GWAS study bias in European and American samples compared to genomic background. All standardizations are computed as the ratio of minor alleles to total alleles per population minus the mean ratio across all individuals from all populations, then all divided by the standard deviation of this ratio. Error bars show the standard error of the mean. A) Standardized across the whole genome. B) Standardized across all sites from the GWAS catalog. C-D) Allele frequencies in local ancestry calls from admixed AMR and AFR/AMR samples are specifically enriched on European tracts and depleted on African and Native American tracts across all genotyped sites and specifically at GWAS sites. Minor allele frequency fraction across C) all sites in admixed AFR/AMR and AMR populations stratified by local ancestry tracts, and D) sites from the GWAS catalog in admixed AFR/AMR and AMR populations stratified by local ancestry tracts.
Figure S9 – Imputation accuracy by population for chromosome 9. A) Illumina OmniExpress. B) Affymetrix Axiom World Array LAT
Figure S10 – Imputation accuracy by population assessed using a leave-one-out strategy, stratified by diploid local ancestry on chromosome 9 for the Affymetrix Axiom World Array LAT genotyping array.
Figure S11 – Standardized polygenic risk score distributions for: A) northern/southern European height, B) asthma, C) waist-hip ratio, D) body mass index, E) Crohn’s disease, and F) ulcerative colitis.
Figure S12 – Simulation runs for the same parameter set ($h^2=0.67$, $m=1000$) and same causal variants with varying effect sizes resulting in a wide range of possible biases in inferred polygenic risk scores across populations.
Figure S13 - Violin plots show Pearson's correlation across 50 iterations per parameter set between true and inferred polygenic risk scores across differing genetic architectures, including \( m=200, 500, \) and \( 1,000 \) causal variants and \( h^2=0.67 \), as in Figure 5. The “ALL” population correlations were performed on population mean-centered true and inferred polygenic risk scores.
Figure S14 – Genetic risk prediction differences across populations. A) Allele frequency versus inferred odds ratio for sites included in inferred polygenic risk scores for each population across 500 simulations, as in Figure 5A-B. B) Log odds ratio by inferred polygenic risk score. The 10,000 individuals with the highest total liability per population were designated as cases, and 10,000 random other individuals in the population were designated as controls. The polygenic risk scores were converted to ordinal deciles, and contrasted with the 1st decile through logistic regression, as: case/control status predicted by polygenic risk decile and population label. Error bars indicate the standard error of the mean across 500 replicates with $h^2=0.67$ and $m=1000$ causal variants, as in Figure 5A-B. Small divergence from the population trend of prediction accuracy across the AFR and EAS are driven by differences in population-specific heritabilities arising from different numbers of population-private causal alleles (i.e. more AFR variants in general gives rise to more AFR-specific causal variants).
**Table S1 – Population names and abbreviations**

<table>
<thead>
<tr>
<th>Population</th>
<th>Code</th>
<th>Super population</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Esan in Nigeria</td>
<td>ESN</td>
<td>AFR</td>
<td>99</td>
</tr>
<tr>
<td>Gambian in Western Division, Mandinka</td>
<td>GWD</td>
<td>AFR</td>
<td>113</td>
</tr>
<tr>
<td>Luhya in Webuye, Kenya</td>
<td>LWK</td>
<td>AFR</td>
<td>99</td>
</tr>
<tr>
<td>Mende in Sierra Leone</td>
<td>MSL</td>
<td>AFR</td>
<td>85</td>
</tr>
<tr>
<td>Yoruba in Ibadan, Nigeria</td>
<td>YRI</td>
<td>AFR</td>
<td>108</td>
</tr>
<tr>
<td>African Caribbean in Barbados</td>
<td>ACB</td>
<td>AFR/AMR</td>
<td>96</td>
</tr>
<tr>
<td>People with African Ancestry in Southwest USA</td>
<td>ASW</td>
<td>AFR/AMR</td>
<td>61</td>
</tr>
<tr>
<td>Colombians in Medellin, Colombia</td>
<td>CLM</td>
<td>AMR</td>
<td>94</td>
</tr>
<tr>
<td>People with Mexican Ancestry in Los Angeles, CA, USA</td>
<td>MXL</td>
<td>AMR</td>
<td>64</td>
</tr>
<tr>
<td>Peruvians in Lima, Peru</td>
<td>PEL</td>
<td>AMR</td>
<td>85</td>
</tr>
<tr>
<td>Puerto Ricans in Puerto Rico</td>
<td>PUR</td>
<td>AMR</td>
<td>104</td>
</tr>
<tr>
<td>Chinese Dai in Xishuangbanna, China</td>
<td>CDX</td>
<td>EAS</td>
<td>93</td>
</tr>
<tr>
<td>Han Chinese in Beijing, China</td>
<td>CDX</td>
<td>EAS</td>
<td>103</td>
</tr>
<tr>
<td>Southern Han Chinese</td>
<td>CHS</td>
<td>EAS</td>
<td>105</td>
</tr>
<tr>
<td>Japanese in Tokyo, Japan</td>
<td>JPT</td>
<td>EAS</td>
<td>104</td>
</tr>
<tr>
<td>Kinh in Ho Chi Minh City, Vietnam</td>
<td>KHV</td>
<td>EAS</td>
<td>99</td>
</tr>
<tr>
<td>Utah residents (CEPH) with Northern and Western European ancestry</td>
<td>CEU</td>
<td>EUR</td>
<td>99</td>
</tr>
<tr>
<td>British in England and Scotland</td>
<td>GBR</td>
<td>EUR</td>
<td>91</td>
</tr>
<tr>
<td>Finnish in Finland</td>
<td>FIN</td>
<td>EUR</td>
<td>99</td>
</tr>
<tr>
<td>Iberian Populations in Spain</td>
<td>IBS</td>
<td>EUR</td>
<td>107</td>
</tr>
<tr>
<td>Toscani in Italia</td>
<td>TSI</td>
<td>EUR</td>
<td>107</td>
</tr>
<tr>
<td>Bengali in Bangladesh</td>
<td>BEB</td>
<td>SAS</td>
<td>86</td>
</tr>
<tr>
<td>Gujarati Indians in Houston, TX, USA</td>
<td>GIH</td>
<td>SAS</td>
<td>103</td>
</tr>
<tr>
<td>Indian Telugu in the UK</td>
<td>ITU</td>
<td>SAS</td>
<td>102</td>
</tr>
<tr>
<td>Punjabi in Lahore, Pakistan</td>
<td>PJL</td>
<td>SAS</td>
<td>96</td>
</tr>
<tr>
<td>Sri Lankan Tamil in the UK</td>
<td>STU</td>
<td>SAS</td>
<td>102</td>
</tr>
</tbody>
</table>
**Table S2** – Three-way admixture proportions between recently admixed populations in the Americas. Values are computed at K=3 on common autosomal SNPs using ADMIXTURE with mean percentages ± standard deviations.

<table>
<thead>
<tr>
<th></th>
<th>AFR</th>
<th>EUR</th>
<th>NAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACB</td>
<td>88.0% (7.7%)</td>
<td>11.7% (7.3%)</td>
<td>0.3% (1.1%)</td>
</tr>
<tr>
<td>ASW</td>
<td>75.6% (13.8%)</td>
<td>21.3% (9.1%)</td>
<td>3.1% (9.2%)</td>
</tr>
<tr>
<td>CLM</td>
<td>7.8% (13.8%)</td>
<td>66.6% (12.8%)</td>
<td>25.7% (9.3%)</td>
</tr>
<tr>
<td>MXL</td>
<td>4.3% (2.2%)</td>
<td>48.7% (18.6%)</td>
<td>47.0% (19.1%)</td>
</tr>
<tr>
<td>PEL</td>
<td>2.5% (5.4%)</td>
<td>20.2% (12.0%)</td>
<td>77.3% (14.2%)</td>
</tr>
<tr>
<td>PUR</td>
<td>13.9% (5.4%)</td>
<td>73.2% (10.0%)</td>
<td>12.9% (3.6%)</td>
</tr>
</tbody>
</table>
Table S3 – Comparison of mean ancestry proportions and ratio on chromosome X versus autosomes across populations. Per Lind et al\textsuperscript{10}, proportion X in a population = (fraction male + 2*fraction female) / 1.5, and proportion autosome in a population = fraction male + fraction female. P-values are from two-sided t-tests on individual ancestries (comparisons are not independent as ancestry proportions must sum to one).

<table>
<thead>
<tr>
<th>Ancestry</th>
<th>ACB</th>
<th>ASW</th>
<th>CLM</th>
<th>MXL</th>
<th>PEL</th>
<th>PUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative X/autosome % change</td>
<td>AFR 4.01</td>
<td>0.83</td>
<td>-2.02</td>
<td>-20.32</td>
<td>50.75</td>
<td>12.69</td>
</tr>
<tr>
<td>EUR -41.73</td>
<td>-17.41</td>
<td>-20.20</td>
<td>-26.60</td>
<td>-41.51</td>
<td>-14.51</td>
<td></td>
</tr>
<tr>
<td>NAT 558.04</td>
<td>87.41</td>
<td>52.70</td>
<td>28.49</td>
<td>9.37</td>
<td>66.89</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>AFR 8.9e-2</td>
<td>7.7e-1</td>
<td>9.8e-1</td>
<td>6.8e-2</td>
<td>3.5e-1</td>
<td>4.1e-1</td>
</tr>
<tr>
<td>EUR 1.0e-3</td>
<td>8.9e-2</td>
<td>1.4e-7</td>
<td>7.9e-4</td>
<td>4.5e-6</td>
<td>1.5e-7</td>
<td></td>
</tr>
<tr>
<td>NAT 7.2e-9</td>
<td>1.1e-1</td>
<td>4.0e-9</td>
<td>3.9e-4</td>
<td>1.3e-3</td>
<td>1.4e-10</td>
<td></td>
</tr>
</tbody>
</table>
Table S4 – Empirical polygenic risk score details. OR = odds ratio

<table>
<thead>
<tr>
<th>Trait</th>
<th>Reference</th>
<th>Effect size</th>
<th>Number of clumps with p≤1e-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>Wood et al, 2014</td>
<td>Beta</td>
<td>35,194</td>
</tr>
<tr>
<td>Female WHR</td>
<td>Shungin et al, 2015</td>
<td>Beta</td>
<td>7,351</td>
</tr>
<tr>
<td>T2D, EUR</td>
<td>Gaulton et al, 2015</td>
<td>Log(OR)</td>
<td>515</td>
</tr>
<tr>
<td>T2D, Multi-ethnic</td>
<td>Mahajan et al, 2014</td>
<td>Log(OR)</td>
<td>11,577</td>
</tr>
<tr>
<td>Asthma</td>
<td>Moffatt et al, 2010</td>
<td>Log(OR)</td>
<td>4,786</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>Ripke et al, 2014</td>
<td>Log(OR)</td>
<td>22,047</td>
</tr>
<tr>
<td>BMI</td>
<td>Locke et al, 2015</td>
<td>Beta</td>
<td>9,445</td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>Jostins et al, 2012</td>
<td>Log(OR)</td>
<td>19,637</td>
</tr>
<tr>
<td>Ulcerative colitis</td>
<td>Jostins et al, 2012</td>
<td>Log(OR)</td>
<td>19,078</td>
</tr>
</tbody>
</table>
References