

# Existence and implications of population variance structure

Shaila Musharoff<sup>1\*</sup>, Danny Park<sup>1,†</sup>, Andy Dahl<sup>1</sup>, Joshua Galanter<sup>1,2,3,‡</sup>, Xuanyao Liu<sup>4</sup>, Scott Huntsman<sup>2</sup>, Celeste Eng<sup>1</sup>, Esteban G. Burchard<sup>1,2</sup>, Julien F. Ayroles<sup>5§</sup>, Noah Zaitlen<sup>1§\*\*</sup>

<sup>1</sup> School of Medicine, University of California, San Francisco, San Francisco, CA, 94143, USA

<sup>2</sup> Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, 94143, USA

<sup>3</sup> Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, 94143, USA

<sup>4</sup> Department of Human Genetics, University of Chicago, Chicago, IL, 60637, USA

<sup>5</sup> Ecology and Evolutionary Biology and Lewis-Sigler for Integrative Genomics, Princeton University, Princeton, NJ, 08544, USA

§ These authors contributed equally to this work

Present address: <sup>†</sup>Encompass Bioscience Inc., San Francisco, CA, 94158, USA

Present address: <sup>‡</sup>Genentech, South San Francisco, 94080, USA

\* For correspondence: [shaila.musharoff@gmail.com](mailto:shaila.musharoff@gmail.com) (SM) or [noah.zaitlen@ucsf.edu](mailto:noah.zaitlen@ucsf.edu) (NZ)

## Abstract

Identifying the genetic and environmental factors underlying phenotypic differences between populations is fundamental to multiple research communities. To date, studies have focused on the relationship between population and phenotypic mean. Here we consider the relationship between population and phenotypic variance, i.e., “population variance structure.” In addition to gene-gene and gene-environment interaction, we show that population variance structure is a direct consequence of natural selection. We develop the ancestry double generalized linear model (ADGLM), a statistical framework to jointly model population mean and variance effects. We apply ADGLM to several deeply phenotyped datasets and observe ancestry-variance associations with 12 of 44 tested traits in ~113K British individuals and 3 of 14 tested traits in ~3K Mexican, Puerto Rican, and African-American individuals. We show through extensive simulations that population variance structure can both bias and reduce the power of genetic association studies, even when principal components or linear mixed models are used. ADGLM corrects this bias and improves power relative to previous methods in both simulated and real datasets. Additionally, ADGLM identifies 17 novel genotype-variance associations across six phenotypes.

## Introduction

Many complex phenotypes differ dramatically in their distributions between populations due to genetic and environmental factors. Both broad<sup>1,2</sup> and fine-scale<sup>3</sup> population differences are central to epidemiology<sup>4</sup>, pharmacogenomics<sup>5,6</sup>, biomedicine<sup>7</sup>, and population genetics<sup>8,9</sup>. In the context of association studies, statistical correction methods for population structure, such as principal components<sup>10</sup> and linear mixed models<sup>11</sup>, have helped identify thousands of loci associated with hundreds of complex traits<sup>12</sup>. This underscores the importance of understanding the causes and consequences of fine-scale population variation.

To date, studies of phenotypic differences between populations and statistical correction methods have primarily focused on variation in population means. As we demonstrate below, while studying fine-scale population structure in UK Biobank, we discovered that phenotypic variance, in addition to phenotypic mean, varies between populations. Such “population variance structure” (in analogy to “population mean structure”) can produce substantial phenotypic differences between populations and has major biological and statistical implications. For example, we recently showed for sex-biased diseases, even a small difference in a disease’s liability variance can double its prevalence between groups<sup>13</sup>. Various evolutionary models<sup>14</sup> also suggest that changes in phenotypic variance allow populations to adapt quickly in response to environmental perturbations<sup>15</sup>.

Although the causes and consequences of phenotypic variance heterogeneity remain poorly understood, several factors could drive population variance structure. First, it can result from non-linear interactions among genotypes (i.e. epistasis). Admixture between genetically diverse populations can disrupt fine-tuned epistatic interactions, increasing phenotypic variance<sup>16,17</sup>. Similarly, gene-environment interactions<sup>18</sup> (GxE) can induce changes in phenotypic variance when environmental exposures differ between populations. Secondly, population variance structure can emerge under additivity. Phenotypic variance itself is a genetically-controlled quantitative trait<sup>19,20</sup>, and as such the frequency of alleles associated with different levels of variability (vQTLs) may differ across populations. Here we also demonstrate, for the first time, that natural selection can directly induce phenotype-variance structure.

To identify and model population variance structure we develop the *Ancestry Double Generalized Linear Model* (ADGLM). ADGLM accommodates arbitrary phenotypic and covariate distributions while accounting for broad- and fine-scale population structure of phenotypic mean as well as variance. Recent work has shown that modeling ancestry-variance effects can reduce

biases of GWAS test statistics<sup>21,22</sup>. However, these methods are limited to modeling binary responses<sup>21</sup> or major population groups<sup>22</sup>. Other studies tested for genotypes associated with phenotypic variance (vQTLs)<sup>23</sup>, but did not model population-variance relationships<sup>24-26</sup>, which generates false-positives when population variance structure exists. We show via extensive simulations that ADGLM reliably detects phenotypic variance structure and is robust to several violations of model assumptions.

To examine the utility of our approach, we first test for population variance structure with ADGLM in several large human datasets. We discover ancestry-variance associations for 12 of 44 tested phenotypes in ~113K UK Biobank British-ancestry individuals and 3 of 14 tested phenotypes in ~3K Mexican, Puerto Rican, and African-American individuals. Additionally, we find 42 ancestry-variance associations in Mexicans of DNA methylation, an epigenetic mark associated with environment<sup>27</sup>, disease phenotypes<sup>28</sup>, and ethnicity<sup>29</sup>. We further illustrate the utility of ADGLM in the context of genetic association mapping and find that relative to linear regression with principal components, modeling population variance structure leads to an increase in power, both in simulated and real datasets. We release ADGLM as open-source R code.

## Material and Methods

### Phenotypic models

For a continuous phenotype  $y$ , we assume

$$y = \sum_s g_s \beta_{g,s} + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n) \quad (\text{Eq. 1})$$

where  $g_s$  is the genotype of the  $s^{\text{th}}$  SNP,  $\beta_{g,s}$  is the genotype's effect size, and errors in  $\varepsilon$  are assumed to be i.i.d. Gaussian. To model binary phenotypes, this model can be modified into a probit model by treating  $y$  as a liability and then thresholding. The main confounder in genetic association studies is population structure<sup>30</sup>. Linear regression with principal components (LR+PC)<sup>10</sup> corrects for this by including the ancestry covariate  $\theta$ :

$$y = \sum_s g_s \beta_{g,s} + \theta \beta_\theta + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n) \quad (\text{Eq. 2})$$

$\theta$  is often a matrix of genetic principal components, but it can contain ancestry admixture fractions or background covariates like age or sex. Linear mixed models (LMM)<sup>31,32</sup> account for

background genetic relatedness by including suitably normalized SNP genotypes,  $Z$ , as a random effect:

$$y = \sum_s g_s \beta_{g,s} + \theta \beta_\theta + Zb + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n) \quad (\text{Eq. 3})$$

We ran LR+PC as ordinary linear regression for continuous traits and probit regression for binary traits. We ran LMM with pylmm<sup>33</sup>, choosing  $Z$  to have centered and scaled columns. Note that this LMM still models each sample as equally variable (modulo inbreeding).

### A statistical model for population variance structure

Population variance structure induces heteroskedasticity<sup>34</sup>, which violates standard linear model assumptions. Recent tests for heteroskedasticity<sup>35,36</sup> or variance effects<sup>24,25,37</sup> are appropriate when their assumptions are met, but they either cannot adjust for ancestry-variance effects, cannot simultaneously account for continuous ancestry and continuous phenotypic distributions<sup>21,22</sup>, or do not scale to UK Biobank sized cohorts. To jointly model phenotypic mean and variance, we develop a framework based on the double generalized linear model<sup>38</sup> (DGLM), which has link functions and covariates for response mean as well as variance. Since we focus on ancestry-phenotype relationships, we call our framework the *Ancestry Double Generalized Linear Model* (ADGLM). The ADGLM uses standard estimates of ancestry ( $\theta$ ), such as fractional ancestry estimates or genetic principal components:

$$y = \sum_s g_s \beta_{g,s} + \theta \beta_\theta + \varepsilon, \quad \varepsilon_i \sim N(0, f(\sigma_\varepsilon^2 + \theta_i \sigma_\theta^2)) \quad (\text{Eq. 4})$$

where  $\varepsilon_i$  is still entry-wise independent and  $f$  is the variance link function, typically the exponential function<sup>39</sup>. Negative variance effects decrease variance, but the exponential function guarantees that total variance is positive. The ADGLM accommodates dichotomous phenotypes via a probit link function or continuous phenotypes, as well as arbitrary phenotypic mean and variance covariates.

### Association testing with the ADGLM

The ADGLM framework enables likelihood ratio tests (LRTs) for ancestry and genetic associations. A 1 degree-of-freedom (df) test for ancestry-variance effect (i.e. population

variance structure,  $\sigma_\theta^2 \neq 0$ ) uses the null model  $H_0$  and alternative model  $H_2$ , while a 1-df LRT for ancestry-mean effect ( $\beta_\theta \neq 0$ ) uses models  $H_1$  and  $H_2$ :

$$\begin{aligned} H_0: y &= \theta \beta_\theta + \varepsilon, & \varepsilon &\sim N(0, \sigma_\varepsilon^2) \\ H_1: y &= \varepsilon, & \varepsilon &\sim N(0, f(\sigma_\varepsilon^2 + \theta \sigma_\theta^2)) \\ H_2: y &= \theta \beta_\theta + \varepsilon, & \varepsilon &\sim N(0, f(\sigma_\varepsilon^2 + \theta \sigma_\theta^2)) \end{aligned}$$

The ADGLM also enables 1-df association tests of mean genetic effect ( $\beta_g \neq 0$ ;  $H_2$  vs.  $H_3$ ) or variance genetic effect for vQTLs ( $\sigma_g^2 \neq 0$ ;  $H_3$  vs.  $H_4$ ), both of which are corrected for population variance structure via  $\theta$  and  $\sigma_\theta^2$ . In addition, a 2-df test for mean genetic and ancestry-variance effects uses models  $H_0$  and  $H_3$ :

$$\begin{aligned} H_3: y &= g \beta_g + \theta \beta_\theta + \varepsilon, & \varepsilon &\sim N(0, f(\sigma_\varepsilon^2 + \theta \sigma_\theta^2)) \\ H_4: y &= g \beta_g + \theta \beta_\theta + \varepsilon, & \varepsilon &\sim N(0, f(\sigma_\varepsilon^2 + \theta \sigma_\theta^2 + g \sigma_g^2)) \end{aligned}$$

We used the R packages “dglm” and “glmx” for continuous and binary phenotypes, respectively, and the exponential variance link function throughout. Though we include a variance intercept term in continuous phenotypic models, we constrain it to one ( $1 = \exp(0)$ ) in binary phenotypic models to obtain identification. In the “dglm” package, standard errors of variance terms are approximated based on the leverages of the variance covariates<sup>40</sup> and thus do not depend on the phenotype if it is scaled. These ancestry and genetic association tests, along with the diagnostic test for residual variance population structure, are implemented in the ADGLM code we released.

### Simulating data from a structured population

We simulated data from a structured sample of two population as follows. Ancestry of the  $i^{th}$  individual,  $\theta_i$ , is 1 for individuals from population 1, and 0 otherwise. We simulated the  $s^{th}$  SNP genotype of the  $i^{th}$  individual from population  $j$  as  $g_{is} \sim \text{Binom}(2, p_{sj})$ , where  $p_{sj}$  is the SNP MAF in population  $j$ . We next simulated independent errors as  $\varepsilon_i \sim N(0, \exp(\sigma_\varepsilon^2 + \theta_i \sigma_\theta^2))$  and phenotypes as  $y_i = \beta_g g_{is} + \theta_i \beta_\theta + \varepsilon_i$ . We took a sample of 200 individuals (100 per population), which has population variance structure when  $\sigma_\theta^2$  is non-zero. For the Figure 2 simulations with no genetic effect,  $m = 10,000$  SNPs,  $\beta_g = 0$ ,  $\beta_\theta = 0$ ,  $\sigma_\varepsilon^2 = 0.2$ ,  $\sigma_\theta^2 = 1.4$ . For the Table 1 simulations,  $m = 10,000$  SNPs,  $\beta_g = 0.8$ ,  $\beta_\theta = 0.2$ ,  $\sigma_\varepsilon^2 = 0.1$ ,  $\sigma_\theta^2 = 1$ . For the Figure 3

simulations with a true genetic effect,  $m = 10,000$  SNPs,  $\beta_g = 0.6$ ,  $\beta_\theta = 0.3$ ,  $\sigma_\varepsilon^2 = 0$ , and  $\sigma_\theta^2$  is one of 35 equally-spaced values between 0 and 2.

### Simulating data from an admixed population

We simulated data from an admixed population composed of two source populations with a given  $F_{ST}$  as follows. We first drew the  $s^{th}$  SNP ancestral minor allele frequency as  $p_s \sim U(0.01, 0.5)$ . We simulated source population MAFs with the Balding-Nichols model<sup>41</sup> as  $p_{sj} \sim \text{Beta}(p_s(1 - F_{st})/F_{st}, (1 - p_s)(1 - F_{st})/F_{st})$ ,  $j = 1, 2$ . The  $i^{th}$  individual's population 1 ancestry fraction was drawn as  $\theta_i \sim U(0.5, 0.9)$ . For each allele ( $k = 1, 2$ ), we drew local ancestries as  $\gamma_k \sim \text{Bin}(1, \theta_i)$  and haploid genotypes as  $l_k \sim \text{Beta}(1, p_{sj})$ , where  $p_{sj}$  is the MAF source population  $\gamma_k$ . We formed the diploid genotype  $g_{is}$  of the  $i^{th}$  individual at SNP  $s$  as the sum of haploid genotypes. Finally, we simulated independent errors as  $\varepsilon_i \sim N(0, \exp(\sigma_\varepsilon^2 + \theta_i \sigma_\theta^2))$  and phenotypes as  $y_i = g_{is} \beta_g + \theta_i \beta_\theta + \varepsilon_i$ .

### Simulating data from an admixed population after differential selection

We simulated data as in “Simulating data from an admixed population” with the following modifications. For values of  $T$  between 0.05 and 0.25 in steps of 0.025, we simulated the  $i^{th}$  individual's ancestry as  $\theta_i \sim N(T, 0.15)$  and truncated it to (0, 1). We drew the  $s^{th}$  SNP effect size as  $\beta_{gs} \sim N(0, 0.2)$ . We then changed effect signs to induce a genetically-based correlation of phenotype and ancestry caused by three strengths of selection. Under neutrality, the sign of  $\beta_{gs}$  is unchanged, so  $\beta_{gs}$  is uncorrelated with ancestry. Under weak selection, the sign of  $\beta_{gs}$  is made positive with probability  $p = \frac{p_{s1}}{p_{s1} + p_{s2}}$  and negative with probability  $1 - p$ , where  $p_{si}$  is population  $i$  MAF. Since the Balding-Nichols model produces identical frequency spectra for all populations,  $p = 0.5$ , and  $\beta_{gs}$  and ancestry are perfectly correlated at half of the SNPs. Finally, under strong selection, the sign of  $\beta_{gs}$  is made positive if  $p_{s1} > p_{s2}$  and negative otherwise, so  $\beta_{gs}$  and ancestry are perfectly correlated at all SNPs. These sign changes result in effect sizes  $\beta_{gs}^*$ . For the  $i^{th}$  individual, we simulated independent error as  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$  and phenotype as  $y_i = \sum_s \beta_{gs}^* g_{is} + \varepsilon_i$ . We did this for 2000 SNPs from a sample of 100 individuals for 1000 replicates.

## Outlier simulations

We simulated data from 1000 individuals for 1000 replicate simulations under the null ( $\sigma_{\theta}^2 = 0$ ) and alternative ( $\sigma_{\theta}^2 = 0.4$ ) of the ancestry-variance test. For simulations with heavy-tailed errors, we simulated errors  $\varepsilon$  from the  $t$  distribution (df=6), simulated ancestry as  $\theta_i \sim N(0.7, 0.4)$  truncated to (0, 1), and formed phenotypes as  $y_i = \theta_i * 0.4 + \varepsilon_i$ . In the simulations with real GALA II ancestry for  $\theta_i$ , we simulated errors as  $\varepsilon_i \sim N(0, \exp(0.1 + \theta_i * \sigma_{\theta}^2))$  and phenotypes as  $y_i = \theta_i * 0.4 + \varepsilon_i$ . We transformed phenotypes or ancestry by inverse-variance quantile-normalizing them or truncating them to remove outliers more than two standard deviations from the mean.

## UK Biobank

We obtained UK Biobank data and restricted our analysis to ~113K British-ancestry individuals. We performed quality control steps as in a previous work<sup>42</sup>, resulting in genetic PCs and continuous phenotypes which are standardized to have mean 0 and standard deviation 1. We additionally quantile-normalized continuous phenotypes. For the variance association test, we adjusted for assessment center, genotype array, sex, age, and PCs 1-10 in the mean. We tested for variance effects (age, sex, PCs1-5) one at a time. The associated traits include ten blood traits, 15 disease traits, body mass index (BMI), blood pressure, educational attainment, basal metabolic rate, two measures of baseline lung function (forced expiratory volume in 1 second, FEV1, and forced vital capacity, FVC), age at menopause, hair pigment, skin pigment, and tanning.

## SAGE II and GALA II datasets

The Study of African Americans, Asthma, Genes & Environments (SAGE II)<sup>43</sup> and Genes-Environment and Admixture in Latino Americans (GALA II)<sup>44</sup> studies are comprised of admixed individuals (ages 8-21). Individuals were deeply phenotyped and genotyped. SAGE consists of 2,013 African Americans. The GALA II study consists of 4427 individuals, of whom 1245 are Mexican and 1785 are Puerto Rican. Genotyping resulted in 482,578 autosomal variants after filtering. We removed related individuals by excluding one of each of a pair of individuals with a REAP<sup>45</sup> coefficient > 0.025, leaving 1160 Mexicans and 1612 Puerto Ricans. For both datasets,



we removed genotypes with MAF < 0.05 and removed SNPs or individuals with more than 5% of genotypes missing. Global ancestry fractions were estimated with the program ADMIXTURE<sup>46</sup> with two ancestral groups (Africans and Europeans) for SAGE and three ancestral groups (Native Americans, Africans, and Europeans) for GALA II.

### **SAGE II and GALA II association testing**

We tested the following phenotypes: asthma; allergy-related disease traits (eczema, hives, rhinitis, rash, and sinusitis); continuous traits (BMI, height); FEV1 and FVC; lung function changes after the first ( $\delta_1$ ) and second ( $\delta_2$ ) albuterol administrations. We also tested two skin pigmentation phenotypes: baseline melanin, the average of right and left body measurements of unexposed areas, and tannability, the difference between baseline and exposed melanin measurements<sup>19</sup>. For ancestry association tests, we included K-1 of K ancestry fractions in the variance model: for Mexicans and Puerto Ricans, we tested for African ancestry and included European ancestry as a variance covariate, and analogously for Native American and African, and African and European ancestry. For African-Americans (K=2), no additional ancestry variance covariate is required. For genetic association tests, we did not thin SNPs for LD nor impute missing phenotypic or covariate measurements. Where noted, genomic control<sup>47</sup> was performed by dividing association test statistics by  $\lambda_{GC}$ . We obtained GWAS associations of tested phenotypes from the NHGRI catalog<sup>48</sup> on April 25, 2018 and thinned it to keep the strongest SNP association per locus, leaving 246 SNPs.

### **GALA II methylation**

We used QC for GALA II methylation data from whole blood as described in Galanter et al.<sup>29</sup>, resulting in batch- and cell type-adjusted methylation at 321,503 autosomal probes. Of the 124 Mexican individuals with methylation measurements, we removed those with outlier Native American ancestry (> 2 s.d. from the mean), leaving 117 individuals. We quantile-normalized methylation values and adjusted for age, sex, ancestry fraction, and asthma case status.



## Results

### Sources of population variance structure

Many studies have explored how genotype-by-environment interactions<sup>18</sup> and epistasis<sup>49-51</sup> may lead to a shift in phenotypic variance as a function of allele frequencies or environmental factors. Here, we consider another possibility: that differential selection between populations causes population variance structure under a purely additive model. To address this question, we simulated admixed populations that experienced differential selection.

We first generated allele frequencies at 2,000 SNPs from two ancestral populations under the Balding-Nichols model<sup>41</sup>. We then simulated effect sizes consistent with natural selection by correlating effect size and allele frequency difference between populations. We used a correlation of 0.0 under neutrality, 0.5 for weak selection, and 1.0 for strong selection. Finally, we simulated phenotypes using an additive model for a sample of 100 two-way admixed individuals composed of these ancestral populations with an average ancestry fraction,  $\theta$ . Under neutrality, neither phenotypic mean nor variance depends on ancestry fraction (Figure 1). However, after either weak or strong selection, both phenotypic mean (Figure 1A) and variance (Figure 1B) depend on ancestry. This demonstrates that differential selection between populations is sufficient to induce population variance structure. In humans, strong, genetically-based ancestry-phenotype correlations are likely due to selection<sup>52</sup>, and may therefore be accompanied by population variance structure.

### Ancestry-variance association tests

We first assessed the performance of the ancestry-variance test ( $\sigma_{\theta}^2 \neq 0$ ) with ADGLM by applying it to simulated data from a structured sample of two populations ( $P_1, P_2$ ) with MAFs  $p_1$  and  $p_2$ . Since the MAF difference ( $p_1 - p_2$ ) determines the genetic variance difference ( $2p_1(1 - p_1) - 2p_2(1 - p_2)$ ) between populations, we considered three types of SNPs based on their MAF in the two populations: SNPs with a MAF difference that is large and negative ( $p_1 = 0.05, p_2 = 0.5$ ), large and positive ( $p_1 = 0.5, p_2 = 0.05$ ) and those with no MAF difference ( $p_1 = 0.5, p_2 = 0.5$ ). We simulated 10,000 SNP genotypes and continuous phenotypes for 100 individuals from each population for a range of  $\beta_g, \beta_{\theta}, \sigma_{\theta}^2$  values. Under the null (no ancestry-variance effect  $\sigma_{\theta}^2 = 0$ ), ADGLM is calibrated with a false positive rate of 0.052 when  $\beta_{\theta} = 0$  and 0.054 when  $\beta_{\theta} = 0.2$  at  $\alpha = 0.05$  (also see Table S1). Under the alternative ( $\sigma_{\theta}^2 = 1$ ), population

$P_1$  has greater phenotypic variance than  $P_2$ , creating population variance structure in their combined sample. Here, ADGLM has power 0.463 when  $\beta_\theta = 0$  and 0.445 when  $\beta_\theta = 0.2$  at  $\alpha = 5e^{-7}$ .

### Effect of population variance structure on genetic association tests

Genome-wide association tests commonly correct for population structure by using linear regression with principal components (LR+PC, Eq. 2) or linear mixed models (LMM<sup>31</sup>, Eq. 3 in Methods). We compared the performance of genetic association tests ( $\beta_g \neq 0$ ) with ADGLM, LR+PC, and LMM applied to data simulated as in "Ancestry-variance association tests". First, we tested for a genetic effect on data simulated under the null ( $\beta_g = 0$ ) with population variance structure, resulting in the quantile-quantile (Q-Q) plots in Figure 2. When population MAFs are equal, LR+PC is calibrated (Figure 2B,  $\lambda_{GC} = 1.01$ ). However, when  $P_1$  MAF is greater than  $P_2$  MAF, LR+PC is inflated (Figure 2A,  $\lambda_{GC} = 1.41$ ); when this MAF relationship is reversed, LR+PC is deflated (Figure 2C,  $\lambda_{GC} = 0.59$ ). By contrast, ADGLM is calibrated for all MAFs: in Figs. 2A, 2B, and 2C,  $\lambda_{GC}$  is 0.98, 1.037, and 1.042, respectively. We also applied a standard LMM with ancestry as a fixed effect and the genetic relationship matrix as a random effect. LMM has the same miscalibration as LR+PC (Figure S1), so we do not consider it further.

Next, we assessed the performance of tests for  $\beta_g \neq 0$  on data simulated with a range of mean genetic ( $\beta_g$ ), mean ancestry ( $\beta_\theta$ ), and ancestry variance ( $\sigma_\theta^2$ ) effects (Table 1, Table S1). When  $\beta_g = 0$  and  $\sigma_\theta^2 = 0$ , linear regression without ancestry adjustment (LR) is calibrated in the absence of population mean or variance structure (rows 1-3). However, LR is miscalibrated if there is population mean structure (row 4) or population variance structure and a MAF difference (rows 6-7). LR+PC and ADGLM perform similarly in the absence of population variance structure: they are calibrated (rows 1-4) and have similar power (row 8), despite ADGLM fitting an additional parameter. When there is population variance structure, ADGLM and LR+PC are calibrated if there is no MAF difference (row 5), whereas only ADGLM is calibrated if there is a MAF difference (rows 6-7). When  $\beta_g \neq 0$  and MAFs are the same, ADGLM is more powerful than LR+PC (rows 9-10). When MAFs differ, LR+PC has less power than ADGLM (row 11) or an elevated false positive rate (row 12).

Finally, we examined the power of genetic association tests ( $\beta_g \neq 0$ ) for varying ancestry-variance effects,  $\sigma_\theta^2$ . Power gains of ADGLM over LR+PC increase with  $\sigma_\theta^2$  when MAFs are the

same (Figure 3B) and when  $P_1$  MAF is less than  $P_2$  MAF (Figure 3C). When this MAF relationship is reversed, LR+PC has false positives, and ADGLM retains its power (Figure 3A). Taken together, these results demonstrate that tests for genetic association with LR+PC are miscalibrated and have false positives or false negatives when there is population variance structure while ADGLM is calibrated and powerful.

### **Diagnostic test for population variance structure**

As we showed above, GWAS performed with standard corrections for population structure may result in biased test statistics in the presence of population variance structure. We developed a test for this bias that can be applied to GWAS summary statistics (Supp. Materials). It regresses association test statistics on the difference of expected genetic variances and tests for a non-zero slope, which occurs when there is residual population variance structure. This diagnostic test is well-powered on test statistics from LR+PC applied to data simulated with population variance structure ( $p=3.1 \times 10^{-26}$ , Figure S2) and is implemented in the ADGLM code repository.

### **Sensitivity of ancestry-variance test to model assumptions**

Double generalized linear models, like most linear models, assume regression errors are normally distributed<sup>23</sup>. We assessed the robustness of testing for  $\sigma_{\theta}^2 \neq 0$  with ADGLM to violations of this assumption. We examined the ability of two transformations to reduce Type 1 errors under model misspecification: inverse-quantile normalization (“normalization”) and outlier removal (“truncation”).

We simulated data under the null with heavy-tailed errors ( $t$ -distribution,  $df=6$ ) and applied ADGLM. Although ADGLM is miscalibrated ( $\lambda_{GC}=1.22$ ,  $FPR=0.079$ ), phenotype truncation ( $\lambda_{GC}=0.95$ ,  $FPR=0.049$ ) or normalization ( $\lambda_{GC}=0.97$ ,  $FPR=0.048$ ) recovers calibration (Figure 4A). We next applied ADGLM to simulated data where 90% of replicates are null and found that relative to the original data (TPR=0.344,  $FPR=0.774$ ), truncation (TPR=0.291,  $FPR=0.049$ ) or normalization (TPR=0.315,  $FPR=0.048$ ) improve both power and false positive rate.

Next, we assessed the robustness of tests for ancestry variance effect ( $\sigma_{\theta}^2 \neq 0$ ) to non-normal ancestry distributions. We simulated phenotypes using ancestry fraction found in the

Gene-Environment studies of Asthma in Hispanic/Latino children (GALA II<sup>43</sup>, Figure S3) as described in Methods. We first used the skewed African ancestry distribution from Puerto Ricans, where 2.0% (35) individuals are ancestry outliers. Applied to data simulated under the null, ADGLM is calibrated ( $\lambda_{GC}=0.991$ , FPR=0.047) and minimally affected by ancestry truncation ( $\lambda_{GC}=1.013$ , FPR=0.048) or normalization ( $\lambda_{GC}=1.025$ , FPR=0.051) in Figure 4B. On data simulated under a mix of the null and alternative, performance is similar for original (TPR=0.067, FPR=0.047), truncated (TPR=0.057, FPR=0.047), and normalized data (TPR=0.076, FPR=0.050). Applied to data simulated with the bell-curved European ancestry from Mexicans with only three ancestry outliers, ADGLM is calibrated under the null ( $\lambda_{GC}=1.00$ , FPR=0.050) and minimally affected by ancestry truncation ( $\lambda_{GC}=1.02$ , FPR=0.050) or normalization ( $\lambda_{GC}=1.00$ , FPR=0.051) in Figure 4C. On data simulated under a mix of the null and alternative, performance is similar for original (TPR=0.087, FPR=0.050), truncated (TPR=0.094, FPR=0.051), and normalized data (TPR=0.105, FPR=0.051). Thus, ancestry distribution transformations improve the performance of ancestry-variance tests, though these ancestry distributions do not cause substantial miscalibration.

### Variance effects in UK Biobank

Individuals from the British Isles have fine-scale population structure which is evident in a large sample<sup>53</sup>. To investigate whether ADGLM can detect fine-scale population variance structure, we applied ADGLM to ~113K British-ancestry, deeply-phenotyped individuals from UK Biobank (UKB, Supp. Materials). We tested binary, ordinal, and quantitative phenotypes (scaled to have mean 0 and variance 1). We included assessment center, genotype array, sex, age, and PCs1-10 as mean effects and tested for population variance structure ( $\sigma_{\theta}^2 \neq 0$ ) with ADGLM. We focus on genetic PCs1-5, which represent geographic population structure in UKB; PC1, specifically, is correlated with a geographic north-south cline<sup>42</sup>.

PC1 is associated (nominal  $p < 0.05$ ) with the phenotypic variance of 17 of 29 tested non-disease traits (Table S2), 12 of which are significant after Bonferroni correction (Table 2). Interestingly, 6 of these 12 associations are only with phenotypic variance, and not mean (absolute correlation of phenotype and PC1  $< 0.01$ ). Corpuscular hemoglobin has the strongest PC1 variance association among continuous traits (Figure S4). In addition, PCs2-5 are associated with the variance of 18 traits (Table S3), representing finer-scale population variance

structure: PC3, which is also correlated with a north-south cline<sup>42</sup>, has the strongest of these variance associations.

We also investigated whether age and sex are associated with phenotypic variance because age varies non-linearly with several phenotypes and different sexes represent different environments<sup>13</sup>. Of the 44 phenotypes tested, 33 have age-variance associations, and 17 have sex-variance associations (Table S2). Overall, population variance structure, age- and sex-variance associations are prevalent in a large sample of British-ancestry individuals.

### Variance effects in admixed populations

For the remainder of this work, we focus on three admixed populations from two asthma and allergy studies: Mexicans and Puerto Ricans from GALA II, and African-Americans from the Study of African Americans, Asthma, Genes, & Environments (SAGE II). We analyzed asthma, allergy-related diseases (eczema, hives, rhinitis, rash, and sinusitis), lung function (FEV1, FVC), change in lung function after the first ( $\delta_1$ ) and second ( $\delta_2$ ) albuterol dose, BMI, height, and skin pigmentation (baseline melanin, tanning). We adjusted phenotypic means for age, sex, and ancestry (African and European ancestry fraction for Mexicans and Puerto Ricans; African ancestry fraction for African-Americans).

Using ADGLM ancestry-variance tests ( $\sigma_\theta^2 \neq 0$ ), we find numerous associations (nominal  $p < 0.05$ ) of ancestry in Figure 5 (also see Tables S4-7), as well as age and sex (Tables S4-7) with phenotypic variance. The ancestry-variance effect sign for a given phenotype is the same across populations except for asthma, which has a negative African variance effect in Puerto Ricans and a positive African variance effect in Mexicans. To test for ancestry-variance heterogeneity, we performed a K-1 df LRT for a population with K ancestry fractions. Of the 14 phenotypes tested in three populations, 4 associations in 3 phenotypes are significant at a Bonferroni-corrected level of 0.05 (which is conservative because the phenotypes are highly correlated): asthma and  $\delta_1$  in Puerto Ricans, and  $\delta_1$  and  $\delta_2$  in African-Americans. In addition, six of the phenotypes have previously-documented ancestry-mean associations which are also detected as mean effects with ADGLM ( $\beta_\theta \neq 0$ ): FEV and FEV1 in Puerto Ricans<sup>54</sup>, asthma in Mexicans and Puerto Ricans<sup>55</sup>, baseline melanin<sup>56,57</sup>,  $\delta_1$ , and  $\delta_2$  in African-Americans<sup>58</sup>.

## ADGLM GWAS in admixed populations

We tested for genetic associations ( $\beta_g \neq 0$ ) of common SNPs ( $MAF > 0.05$ ) in the admixed datasets above using both ADGLM, which corrects for population variance structure, and LR+PC, which does not. We represent ancestry using two ancestry fractions (African and European) for Puerto Ricans and Mexicans, and one (African) for African-Americans. Effect sizes for dichotomous traits (such as eczema) cannot be compared directly because they were obtained through probit regression. We discover two novel SNP associations with ADGLM, neither of which is significant with LR+PC: in Table S8, rs9808780 is associated with eczema in Mexicans and Puerto Ricans ( $p = 1.53e-8$ ) and rs113736578 is associated with rash in Puerto Ricans ( $p = 2.14e-8$ ). We next compared ADGLM to LR+PC at GWAS associations in the NHGRI catalog<sup>48</sup>, thinned to one SNP per locus. Since ADGLM is less inflated than LR+PC (Table S9), we genomic control<sup>47</sup> adjusted test statistics when  $\lambda_{GC} \geq 1$  to be maximally conservative. For the 46 GWAS SNPs in our datasets, 12 SNPs replicate with either test ( $p_{adj} < 0.05$ , Table 3;). Of these, 11 have a more significant p-value from ADGLM than LR+PC, indicating that ADGLM has better power to detect genetic associations than LR+PC.

## Variance QTL

We next tested for genetic variance associations ( $\sigma_g^2 \neq 0$ ) with ADGLM to find variance quantitative trait loci (vQTLs) in the admixed datasets above. For Mexicans and Puerto Ricans, we included genotype in the mean model and adjusted for ancestry, age, and sex in the mean and variance models; we did the same for African-Americans, adjusting for significant variance covariates (Table S6). We detect 17 vQTLs after genomic control adjustment ( $p_{adj} < 5e-8$ ) in Table 4; the corresponding  $\lambda_{GC}$  values are in Table S10. The associations with hives in Mexicans, height in Puerto Ricans, asthma in African Americans, and tanning in African Americans are each detected in only one population. Of the 17 genetic variance associations, only 3 also have significant mean effects (rs1640275, rs117344403, rs55837614).

## Methylation association studies

DNA methylation, an epigenetic mark which is affected by environmental factors<sup>27</sup>, varies across disease phenotypes<sup>28</sup> and ancestry<sup>29</sup>. To characterize the relationship of methylation and ancestry variance, we analyzed quantile-normalized methylation from 117 Mexican individuals.

We adjusted for the mean effect of age, sex, ancestry, and asthma case status. We tested for ancestry-mean effects ( $\beta_{\theta} \neq 0$ ) with ADGLM and LR+PC, as well as ancestry-variance effects ( $\sigma_{\theta}^2 \neq 0$ ) with ADGLM, resulting in Q-Q plots in Figure S5 and Manhattan plots in Figure S6. After Bonferroni correction, ADGLM identifies eight loci with ancestry-mean effects and 42 loci with ancestry-variance effects, 4 of which also have significant mean effects (Table S11). LR+PC, by contrast, only identifies one mean association, which is declared as a significant variance association, but not a mean association, by ADGLM.

## Discussion

In this study, we describe the presence of and discuss the importance of population variance structure, the difference of phenotypic variance by population. To model ancestry-variance relationships, we developed a novel statistical framework, the ancestry double generalized linear model (ADGLM). Unlike existing variance models, ADGLM accounts for continuous and discrete definitions of ancestry, arbitrary covariates, and binary or continuous phenotypes. We used ADGLM to discover many ancestry-variance associations in a British-ancestry and admixed human populations for a wide range of binary and continuous traits, including diseases and methylation, many of which have been subject to natural selection<sup>42,56,57,59,60</sup>.

When ancestry is related to phenotypic variance, genetic association tests with standard population structure corrections (e.g. linear regression with principal components adjustment or linear mixed models) are miscalibrated as a function of minor allele frequency. This miscalibration has been observed for binary traits and can be attributed to the inability of standard LMM to model differences in disease prevalence<sup>21</sup>. We additionally observed this miscalibration for continuous traits and showed that it is a consequence of unmodeled population variance structure. Though not always apparent in a genome-wide Q-Q plot, this miscalibration can be readily detected by our diagnostic test which operates on summary statistics. ADGLM addresses these problems and association tests with ADGLM are both calibrated and well-powered for simulated and empirical human data.

The numerous variance associations we observed imply that previously conducted GWAS using LR+PC or LMM have residual population variance structure. The impact can be substantial as demonstrated by the inflation of height GWAS test statistics in LDscore analysis<sup>61</sup>. Two recent studies<sup>62,63</sup> also found evidence for incomplete population structure



correction in large cohort studies, including UK Biobank; based on our analysis, this may be due in part to unmodeled population variance structure. If phenotypes and principal components are available, population variance structure can be detected with ADGLM. Though the association of the square of a centered, scaled phenotype with principal components implies population variance structure may exist<sup>64</sup>, it is not a direct test.

In addition to acting as a statistical confounder, population variance structure has important biological implications, including in medical genetics<sup>64</sup>. Intuitively, differences in phenotypic distribution between populations imply that the fraction of individuals in the phenotypic tails differs between populations, and as such, longer tails may indicate a greater disease burden. As we previously showed, small differences in phenotypic variance between sexes can create large differences in disease liability<sup>13</sup>. Here, we estimate different asthma African ancestry-variance effects for Mexicans ( $17.2 \pm 7.6$ ) and Puerto Ricans ( $-4.05 \pm 2.9$ ). Mexicans and Puerto Ricans living in the U.S. differ dramatically in their asthma prevalence (8% vs. 22%), which has been referred to as the “Hispanic Paradox”<sup>65</sup>. These ancestry-variance associations might partially explain this difference.

In the 1940s, Waddington proposed that phenotypic variability is under genetic control, biological systems evolve to maintain homeostasis under a certain range of environmental or genetic perturbations, and that changing the environment to be outside the normal range will shift an optimum that has been shaped by stabilizing selection over many generations<sup>66</sup>. Under this model, increasing admixture proportion or shifting environmental conditions will lead to an increase in population variance, a process known as decanalization<sup>15</sup>. Although this phenomenon is well-documented in other species, it has seldom been described in humans, where it has been proposed as an explanation for the dramatic increase in non-communicable complex disease prevalence<sup>67</sup>. The ability of ADGLM to identify population variance structure as a function of admixture proportion or specific environmental context offers a new avenue to identify factors associated with variance heterogeneity, and thus potential drivers of decanalization.

ADGLM could be used as a screening tool for signals consistent with the presence of epistatic or genotype-by-environment interactions. ADGLM vQTL, which are controlled for variance population structure, may point to unmodeled interactions (either GxG or GxE)<sup>68</sup> that can then be tested for interactions with other loci or specific environmental variables<sup>49</sup>. Additionally, the presence of population variance structure may affect admixture mapping

efforts<sup>69</sup> and could be corrected for with ADGLM. Finally, we could reduce the running time of the ADGLM GWAS roughly to ordinary GWAS by fitting the background variance components only once<sup>70</sup>, rather than once per SNP.

In conclusion, we find pervasive population variance structure in multiple human populations. As human studies increase in size and diversity, models that account for population variance structure, such as ADGLM, will be required for interpretable association testing. ADGLM has utility in studies of non-human model systems and natural populations, which have differences in phenotypic variability among groups and variance effects. By focusing primarily on the effect of genetic variation on phenotypic mean and ignoring its effect on variance, we have been missing an important axis contributing to phenotypic variation and disease emergence. Modeling phenotypic variance with ADGLM will enable discoveries along this axis.

## Acknowledgments

We thank Simon Forsberg for data analysis support and Joel Mefford for insightful discussions. This research was conducted using the UK Biobank Resource under Application #30397. N.Z., S.M, A.D, and D.P were funded from NIH grants U011U01HG009080, 5R01HG006399, and 5K25HL121295. J.F.A was funded by NIH grant 5R35GM124881. This work was supported in part by the Sandler Family Foundation, the American Asthma Foundation, the RWJF Amos Medical Faculty Development Program, Harry Wm. and Diana V. Hind Distinguished Professor in Pharmaceutical Sciences II, National Institutes of Health R01HL117004, R01HL128439, R01HL135156, 1X01HL134589, R01HL141992, National Institute of Health and Environmental Health Sciences R01ES015794, R21ES24844, the National Institute on Minority Health and Health Disparities P60MD006902 RL5GM118984, R01MD010443 and the Tobacco-Related Disease Research Program under Award Number 24RT-0025, 27IR-0030.

## Web Resources

The URLs for data presented herein are as follows:

ADGLM, <https://github.com/shailam/adglm>

pylmm, <https://github.com/nickFurlotte/pylmm>

## References

1. Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., et al. (2008). Genes mirror geography within Europe. *Nature* 456, 98–101.
2. The 1000 Genomes Project Consortium, Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J.G., Zhu, Y., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
3. Moreno-Estrada, A., Gignoux, C.R., Fernandez-Lopez, J.C., Zakharia, F., Sikora, M., Contreras, A.V., Acuña-Alonzo, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., et al. (2014). Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* 344, 1280–1285.
4. Marshall, M.C. (2005). Diabetes in African Americans. *Postgraduate Medical Journal* 81, 734–740.
5. Wood, A.J.J. (2001). Racial Differences in the Response to Drugs — Pointers to Genetic Differences. *N Engl J Med* 344, 1394–1396.
6. A Perera, M., PharmD, L.H.C., PharmD, N.A.L., MS, E.R.G., MS, A.K., BS, R.D., PhD, A.P., PhD, D.C.C., BS, J.W., PhD, N.L., et al. (2013). Genetic variants associated with warfarin dose in African- American individuals: a genome-wide association study. *The Lancet* 382, 790–796.
7. Burchard, E.G. (2006). Importance of Race/Ethnicity and Genetics in Biomedical Research and Clinical Practice: Lessons Learned from the Genetics of Asthma in Latino Americans (Gala) Study. *American Journal of Epidemiology* 163, S84–S84.
8. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–496.
9. Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. (2014). A genetic atlas of human admixture history. *Science* 343, 747–751.
10. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 2010 42:7 38, 904–909.
11. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics* 88, 76–82.
12. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics* 101, 5–22.
13. Traglia, M., Bseiso, D., Gusev, A., Adviento, B., Park, D.S., Mefford, J.A., Zaitlen, N., and Weiss, L.A. (2017). Genetic Mechanisms Leading to Sex Differences Across Common Diseases and Anthropometric Traits. *Genetics* 205, 979–992.

14. Philippi, T., and Seger, J. (1989). Hedging one's evolutionary bets, revisited. *Trends in Ecology & Evolution* 4, 41–44.
15. Gibson, G., and Wagner, G. (2000). Canalization in evolutionary genetics: a stabilizing theory? *Bioessays* 22, 372–380.
16. Ackermann, R.R. (2011). Phenotypic traits of primate hybrids: Recognizing admixture in the fossil record. *Evolutionary Anthropology: Issues, News, and Reviews* 19, 258–270.
17. Pélabon, C., Hansen, T.F., Carter, A.J.R., and Houle, D. (2010). Evolution of variation and variability under fluctuating, stabilizing, and disruptive selection. *Evolution* 64, 1912–1925.
18. Murcay, C.E., Lewinger, J.P., and Gauderman, W.J. (2008). Gene-Environment Interaction in Genome-Wide Association Studies. *American Journal of Epidemiology* 169, 219–226.
19. Hill, W.G., and Mulder, H.A. (2010). Genetic analysis of environmental variation. *Genet. Res.* 92, 381–395.
20. Ayroles, J.F., Buchanan, S.M., O'Leary, C., Skutt-Kakaria, K., Grenier, J.K., Clark, A.G., Hartl, D.L., and de Bivort, B.L. (2015). Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proceedings of the National Academy of Sciences* 112, 6706–6711.
21. Chen, H., Wang, C., Conomos, M.P., Stilp, A.M., Li, Z., Sofer, T., Szpiro, A.A., Chen, W., Brehm, J.M., Celedón, J.C., et al. (2016). Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *The American Journal of Human Genetics* 98, 653–666.
22. Conomos, M.P., Laurie, C.A., Stilp, A.M., Gogarten, S.M., McHugh, C.P., Nelson, S.C., Sofer, T., Fernández-Rhodes, L., Justice, A.E., Graff, M., et al. (2016). Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *The American Journal of Human Genetics* 98, 165–184.
23. Rönnegård, L., Felleki, M., Fikse, F., Mulder, H.A., and Strandberg, E. (2010). Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. *Genet Sel Evol* 42, 8–10.
24. Corty, R.W., Kumar, V., Tarantino, L., Takahashi, J., and Valdar, W. (2018). Mean-Variance QTL Mapping Identifies Novel QTL for Circadian Activity and Exploratory Behavior in Mice. 1–16.
25. Dumitrascu, B., Darnell, G., Ayroles, J., and Engelhardt, B.E. (2018). Statistical tests for detecting variance effects in quantitative trait studies. *Bioinformatics* 4, e1000049.
26. Cao, Y., Wei, P., Bailey, M., Kauwe, J.S.K., and Maxwell, T.J. (2014). A versatile omnibus test for detecting mean and variance heterogeneity. *Genet. Epidemiol.* 38, 51–59.
27. Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* 13, 97–109.
28. Torfi, Y., Bitarafan, N., and Rajabi, M. (2015). Impact of socioeconomic and environmental factors on atopic eczema and allergic rhinitis: a cross sectional study. *Excli J* 14, 1040–1048.

29. Galanter, J.M., Gignoux, C.R., Oh, S.S., Torgerson, D., Pino-Yanes, M., Thakur, N., Eng, C., Hu, D., Huntsman, S., Farber, H.J., et al. (2017). Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *Elife* 6, 1655.
30. Marchini, J., Cardon, L.R., Phillips, M.S., and Donnelly, P. (2004). The effects of human population structure on large genetic association studies. *Nature Genetics* 2010 42:7 36, 512–517.
31. Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J., and Eskin, E. (2008). Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics* 178, 1709–1723.
32. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics* 88, 76–82.
33. Sul, J.H., Bilow, M., Yang, W.-Y., Kostem, E., Furlotte, N., He, D., and Eskin, E. (2016). Accounting for Population Structure in Gene-by-Environment Interactions in Genome-Wide Association Studies Using Mixed Models. *PLoS Genet* 12, e1005849.
34. Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. (Berkeley, Calif.: University of California Press), pp. 221–233.
35. Schultz, B.B. (1985). Levene's Test for Relative Variation. *Systematic Biology* 34, 449–456.
36. Brown, M.B., and Forsythe, A.B. (1974). The Small Sample Behavior of Some Statistics Which Test the Equality of Several Means. *Technometrics* 16, 129–132.
37. Cao, Y., Maxwell, T.J., and Wei, P. (2014). A Family-Based Joint Test for Mean and Variance Heterogeneity for Quantitative Traits. *Annals of Human Genetics* 79, 46–56.
38. Smyth, G.K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society Series* 50, 47–60.
39. Corty, R.W., and Valdar, W. (2018). vqtl: An R package for Mean-Variance QTL Mapping. *bioRxiv* 1–7.
40. Smyth, G.K., and Verbyla, A.P. (1999). Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics* 10, 695–709.
41. Balding, D.J., and Nichols, R.A. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* 96, 3–12.
42. Galinsky, K.J., Loh, P.-R., Mallick, S., Patterson, N.J., and Price, A.L. (2016). Population Structure of UK Biobank and Ancient Eurasians Reveals Adaptation at Genes Influencing Blood Pressure. *The American Journal of Human Genetics* 99, 1130–1139.
43. Borrell, L.N., Nguyen, E.A., Roth, L.A., Oh, S.S., Tcheurekdjian, H., Sen, S., Davis, A., Farber, H.J., Avila, P.C., Brigino-Buenaventura, E., et al. (2013). Childhood Obesity and Asthma Control in the GALA II and SAGE II Studies. *Am J Respir Crit Care Med* 187, 697–702.

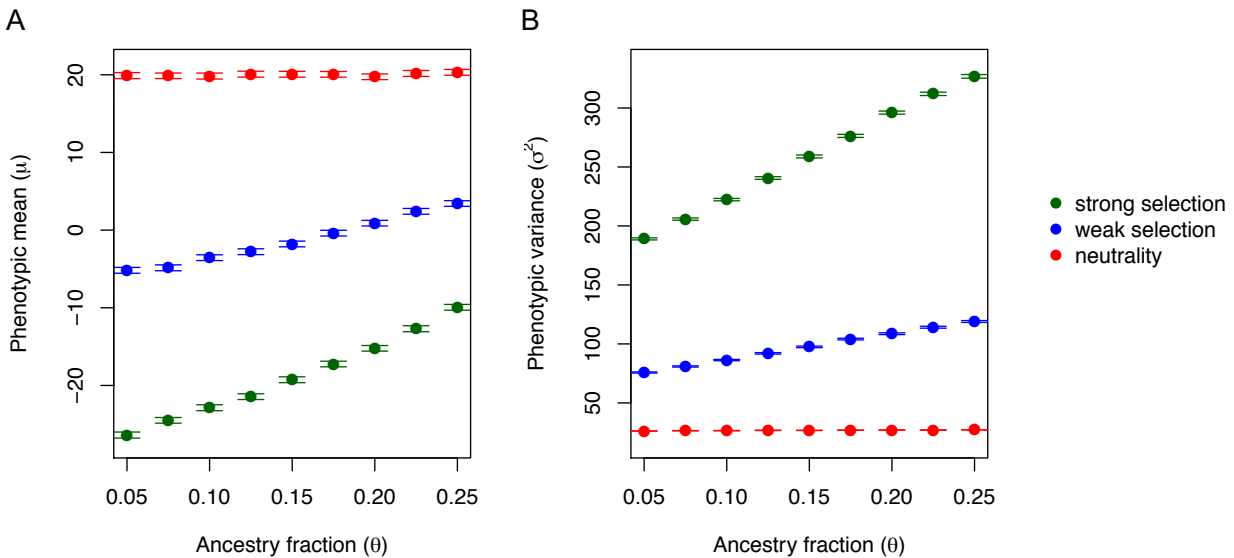
44. Nishimura, K.K., Galanter, J.M., Roth, L.A., Oh, S.S., Thakur, N., Nguyen, E.A., Thyne, S., Farber, H.J., Serebrisky, D., Kumar, R., et al. (2013). Early-Life Air Pollution and Asthma Risk in Minority Children. The GALA II and SAGE II Studies. *Am J Respir Crit Care Med* 188, 309–318.
45. Thornton, T., Tang, H., Hoffmann, T.J., Ochs-Balcom, H.M., Caan, B.J., and Risch, N. (2012). Estimating Kinship in Admixed Populations. *Am. J. Hum. Genet.* 91, 122–138.
46. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19, 1655–1664.
47. Devlin, B., and Roeder, K. (1999). Genomic Control for Association Studies. *Biometrics* 55, 997–1004.
48. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* 45, D896–D901.
49. Brown, A.A., Buil, A., Viñuela, A., Lappalainen, T., Zheng, H.-F., Richards, J.B., Small, K.S., Spector, T.D., Dermitzakis, E.T., and Durbin, R. (2014). Genetic interactions affecting human gene expression identified by variance association mapping. *Elife* 3, 1198–16.
50. Paré, G., Cook, N.R., Ridker, P.M., and Chasman, D.I. (2010). On the Use of Variance per Genotype as a Tool to Identify Quantitative Trait Interaction Effects: A Report from the Women's Genome Health Study. *PLoS Genet* 6, e1000981–10.
51. Cordell, H.J. (2002). Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* 11, 2463–2468.
52. Edge, and Rosenberg (2015). A General Model of the Relationship between the Apportionment of Human Genetic Diversity and the Apportionment of Human Phenotypic Diversity. *Human Biology* 87, 313–326.
53. Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Wellcome Trust Case Control Consortium 2, et al. (2015). The fine-scale genetic structure of the British population. *Nature* 519, 309–314.
54. Brehm, J.M., Acosta-Pérez, E., Klei, L., Roeder, K., Barmada, M.M., Boutaoui, N., Forno, E., Cloutier, M.M., Datta, S., Kelly, R., et al. (2012). African ancestry and lung function in Puerto Rican children. *Journal of Allergy and Clinical Immunology* 129, 1484–1490.e1486.
55. Pino-Yanes, M., Thakur, N., Gignoux, C.R., Galanter, J.M., Roth, L.A., Eng, C., Nishimura, K.K., Oh, S.S., Vora, H., Huntsman, S., et al. (2015). Genetic ancestry influences asthma susceptibility and lung function among Latinos. *J. Allergy Clin. Immunol.* 135, 228–235.
56. Martin, A.R., Lin, M., Granka, J.M., Myrick, J.W., Liu, X., Sockell, A., Atkinson, E.G., Werely, C.J., Möller, M., Sandhu, M.S., et al. (2017). An Unexpectedly Complex Architecture for Skin Pigmentation in Africans. *Cell* 171, 1340–1353.e14.
57. Crawford, N.G., Kelly, D.E., Hansen, M.E.B., Beltrame, M.H., Fan, S., Bowman, S.L., Jewett, E., Ranciaro, A., Thompson, S., Lo, Y., et al. (2017). Loci associated with skin pigmentation identified in African populations. *Science* 358, eaan8433.



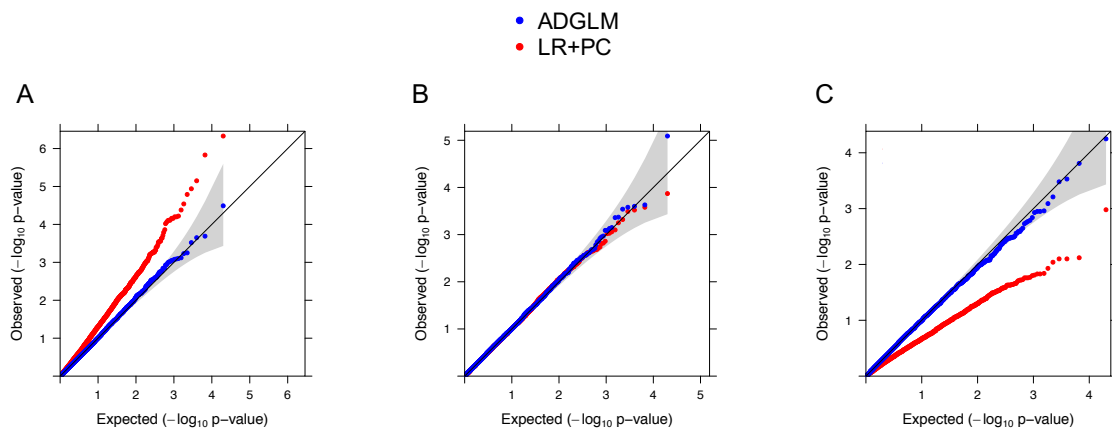
58. Mak, A.C.Y., White, M.J., Eckalbar, W.L., Szpiech, Z.A., Oh, S.S., Pino-Yanes, M., Hu, D., Goddard, P., Huntsman, S., Galanter, J., et al. (2018). Whole-Genome Sequencing of Pharmacogenetic Drug Response in Racially Diverse Children with Asthma. *Am J Respir Crit Care Med* 197, 1552–1564.
59. Zeng, J., Vlaming, R., Wu, Y., Robinson, M.R., Lloyd-Jones, L.R., Yengo, L., Yap, C.X., Xue, A., Sidorenko, J., McRae, A.F., et al. (2018). Signatures of negative selection in the genetic architecture of human complex traits. *Nature Publishing Group* 1–12.
60. Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S.E., Hu, F.B., Duffy, D.L., Zhao, Z.Z., et al. (2008). A Genome-Wide Association Study Identifies Novel Alleles Associated with Hair Color and Skin Pigmentation. *PLoS Genet* 4, e1000074–11.
61. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Publishing Group* 47, 291–295.
62. Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C., Chiang, C.W.K., Hirschhorn, J.N., Daly, M., Patterson, N., et al. (2018). Signals of polygenic adaptation on height have been overestimated due to uncorrected population structure in genome-wide association studies. 1–12.
63. Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., Boyle, E.A., Zhang, X., Racimo, F., Pritchard, J.K., et al. (2018). Reduced signal for polygenic adaptation of height in UK Biobank.
64. Yang, J., Loos, R.J.F., Powell, J.E., Medland, S.E., Speliotes, E.K., Chasman, D.I., Rose, L.M., Thorleifsson, G., Steinthorsdottir, V., Mägi, R., et al. (2012). FTO genotype is associated with phenotypic variability of body mass index. *Nature* 490, 267–272.
65. Franzini, L., Ribble, J.C., and Keddie, A.M. (2001). Understanding the Hispanic paradox. *Ethnicity and Disease* 11, 496–518.
66. WADDINGTON, C.H. (1942). CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS. *Nature* 150, 563–565.
67. Gibson, G. (2009). Decanalization and the origin of complex disease. *Nat Rev Genet* 10, 134–140.
68. Forsberg, S.K.G., and Carlborg, Ö. (2017). On the relationship between epistasis and genetic variance heterogeneity. *J. Exp. Bot.* 68, 5431–5438.
69. Winkler, C.A., Nelson, G.W., and Smith, M.W. (2010). Admixture mapping comes of age. *Annu. Rev. Genom. Hum. Genet.* 11, 65–89.
70. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Publishing Group* 42, 348–354.



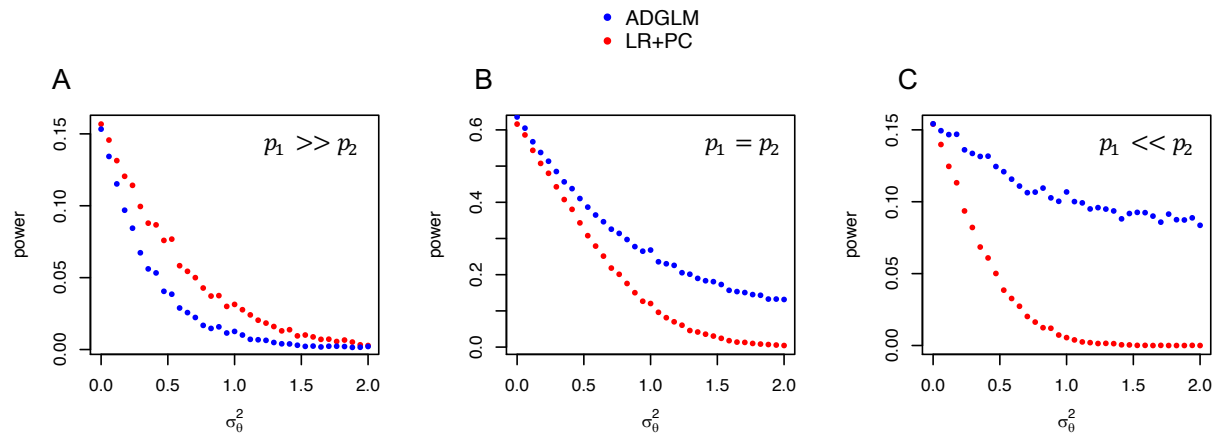
## Figures



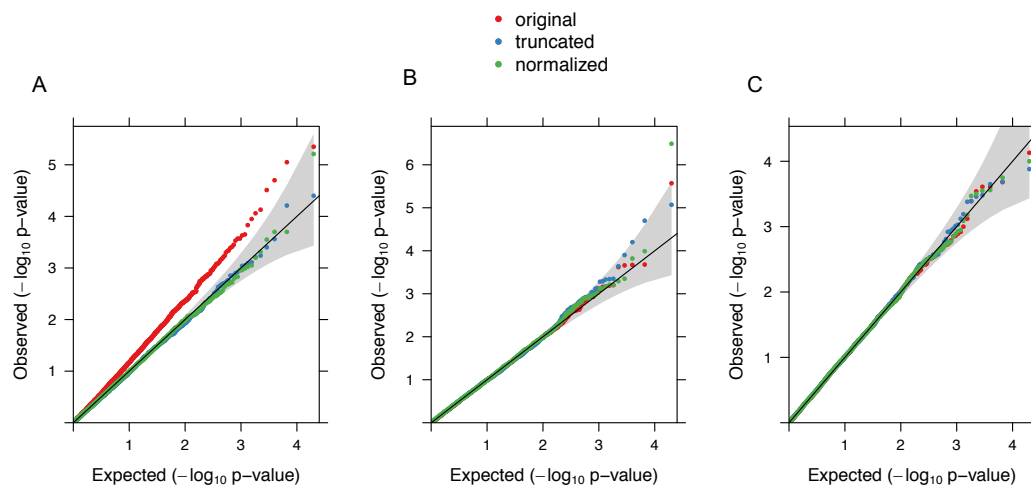
**Figure 1: Differential selection induces population variance structure.** (A) Phenotypic mean,  $\mu$ , and (B) phenotypic variance,  $\sigma^2$ , of an admixed population with average ancestry fraction,  $\theta$ , composed of two populations that experienced neutrality (red), weak selection (blue), or strong selection (green). Points are averages across 1000 simulations and bars denote 95% confidence intervals. Selection induces a dependence of phenotypic mean, as well as phenotypic variance, on ancestry.



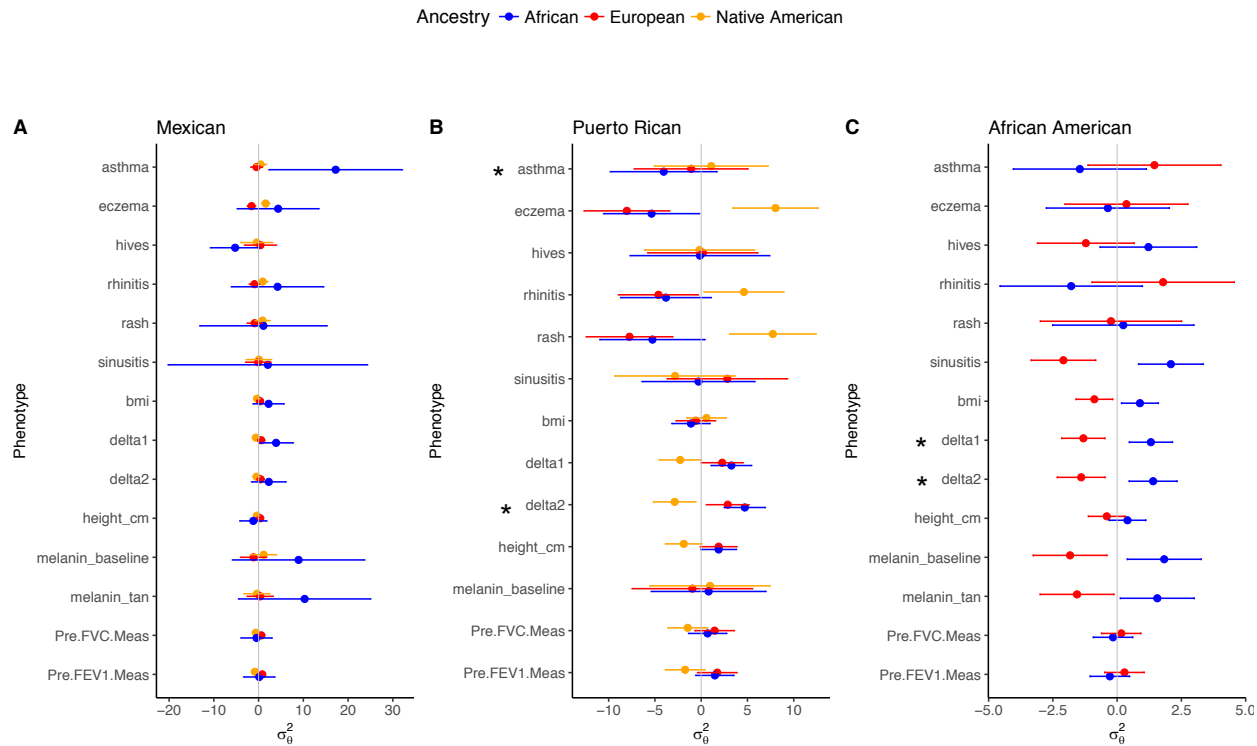
**Figure 2: Q-Q plots of genetic association tests applied simulated null data.** Data simulated with  $\beta_g=0$  and population variance structure for population MAFs: (A)  $p_1 = 0.5, p_2 = 0.05$ , (B)  $p_1 = 0.5, p_2 = 0.5$ , or (C)  $p_1 = 0.05, p_2 = 0.5$ . The null expectation is denoted by the black line and the null 95% confidence interval by the gray band. LR+PC  $-\log_{10}$ (p-values) in red are (A) inflated or (C) deflated when population MAFs differ; those from ADGLM in blue are calibrated.



**Figure 3: Power of genetic association tests applied to simulated data.** Power ( $\alpha = 5e^{-8}$ ) of tests applied to data simulated with varying  $\sigma_\theta^2$  for three MAF cases. (A) For  $p_1 = 0.5$ ,  $p_2 = 0.05$ , LR+PC is enriched for false positives, and ADGLM is well-powered. For (B)  $p_1 = 0.5$ ,  $p_2 = 0.5$  and (C)  $p_1 = 0.05$ ,  $p_2 = 0.5$ , ADGLM has more power than LR+PC when there is population variance structure ( $\sigma_\theta^2 > 0$ ).



**Figure 4: Effect of phenotype and ancestry outliers on ancestry-variance tests.** Q-Q plots of ancestry-variance association tests applied to data simulated under the null ( $\sigma_\theta^2 = 0$ ). (A) Phenotype transformation reduces false positives of data simulated with heavy-tail, t-distributed error. (b, c) Tests applied to data simulated with real (B) African ancestry from Puerto Ricans or (C) European ancestry from Mexicans are calibrated and unaffected by ancestry transformation.



**Figure 5: Ancestry-variance associations in admixed populations.** Associations of African, European, and Native American ancestry with phenotypic variance in (A) Mexicans, (B) Puerto Ricans, and (C) African Americans. Points indicate estimates of ancestry variance effects, bands represent 95% confidence intervals, and starred phenotypes have significant ancestry-variance heterogeneity after Bonferroni correction ( $p < 0.05$ ). The gray vertical line denotes  $\sigma_{\theta}^2 = 0$ .

## Tables

Row	Parameters					Discovery rate		
	$p_1$	$p_2$	$\beta_G$	$\beta_\theta$	$\sigma_\theta^2$	LR	LR+PC	ADGLM
1	0.5	0.5	0	0	0	0.051	0.051	0.054
2	0.05	0.5	0	0	0	0.054	0.052	0.054
3	0.5	0.05	0	0	0	0.054	0.050	0.050
4	0.05	0.5	0	0.2	0	0.140	0.053	0.054
5	0.5	0.5	0	0	1	0.051	0.051	0.055
6	0.5	0.05	0	0	1	0.074	0.086	0.053
7	0.05	0.5	0	0	1	0.032	0.021	0.057
8	0.05	0.5	0.8	0	0	0.993	0.783	0.725
9	0.5	0.5	0.8	0	1	0.698	0.693	0.848
10	0.5	0.5	0.8	0.2	1	0.702	0.702	0.850
11	0.05	0.5	0.8	0.2	1	0.467	0.205	0.606
12	0.5	0.05	0.8	0.2	1	0.838	0.269	0.135

**Table 1: Performance of genetic association tests applied to simulated data.** We report false positive rate ( $\alpha = 0.05$ , rows 1-7) or power ( $\alpha = 5e^{-7}$ , rows 8-12) of tests of genetic effect ( $\beta_g \neq 0$ ) for a range of MAFs ( $p_1, p_2$ ) and genetic ( $\beta_g$ ), ancestry-mean ( $\beta_\theta$ ), and ancestry-variance ( $\sigma_\theta^2$ ) effects. The null hypothesis is true above the double line and false below it. Linear regression (LR) is miscalibrated in the presence of population structure. ADGLM is calibrated and powerful while linear regression with principal components (LR+PC) is often biased or underpowered.

Phenotype	Effect size ( $\sigma_\theta^2$ )	P-value
Corpuscular hemoglobin	-0.025 ± 2.5E-03	7.75E-25
Platelet volume *	0.015 ± 2.5E-03	1.36E-09
Platelet count	-0.012 ± 2.5E-03	7.13E-07
Platelet distribution width *	0.015 ± 2.5E-03	2.14E-09
Erythrocyte (red blood cell) distribution width	-0.019 ± 2.5E-03	3.77E-14
Leukocyte (white blood cell) count	-0.018 ± 2.5E-03	1.03E-13
Educational attainment *	-0.020 ± 2.4E-03	5.73E-16
Basal metabolic rate (z)	-0.024 ± 2.5E-03	3.41E-23
Forced vital capacity (z) *	0.012 ± 2.7E-03	8.89E-06
Hair pigment	-0.023 ± 2.46E-03	1.28E-21
Skin pigment *	-0.074 ± 2.45E-03	9.25E-199
Skin tanning *	0.068 ± 2.45E-03	4.02E-169

**Table 2: UK Biobank variance associations.** Ancestry variance effect sizes estimates, standard errors, and p-values of associations of PC1 with phenotypic variance in UKB British-ancestry individuals. Associations are significant at a threshold of 0.05 after Bonferroni correction. Phenotypes followed by "(z)" are z-scores and starred phenotypes have an absolute correlated with PC1 that is greater than 0.01.

Population	Trait	SNP	LR+PC		ADGLM	
			Effect size ( $\beta_g$ )	P-value	Effect size ( $\beta_g$ )	P-value
MX	Eczema	rs17389644 *	0.252 ± 1.51E-01	1.10E-01	4.90E-4 ± 2.15E-03	3.49E-02
MX	Eczema	rs4796793 *	-0.412 ± 9.53E-02	3.85E-05	-0.0074 ± 4.05E-02	2.76E-05
MX	Melanin	rs16891982 *	0.487 ± 1.70E-01	8.90E-03	0.591 ± 1.58E-01	8.47E-04
MX	Tanning	rs12210050 *	0.734 ± 3.32E-01	4.08E-02	0.748 ± 3.35E-01	2.50E-02
MX	Tanning	rs7969151 *	0.613 ± 2.70E-01	3.63E-02	0.803 ± 2.38E-01	4.10E-03
MX	Tanning	rs154659 *	-0.393 ± 1.92E-01	5.73E-02	-0.397 ± 1.78E-01	2.41E-02
MX	FVC	rs9791644	-0.148 ± 6.33E-02	2.65E-02	-0.131 ± 6.32E-02	4.94E-02
MX	Rhinitis	rs868688 *	-0.137 ± 6.22E-02	3.48E-02	-0.002 ± 8.13E-03	2.25E-02
PR	Eczema	rs13015714 *	-0.14 ± 7.29E-02	6.55E-02	-3.22E-4 ± 7.52E-4	4.15E-03
PR	Melanin	rs1834640 *	-0.374 ± 9.37E-02	1.81E-04	-0.375 ± 9.32E-02	1.41E-04
PR	Melanin	rs6142102 *	0.244 ± 8.55E-02	7.02E-03	0.253 ± 8.40E-02	4.45E-03
PR	Melanin	rs2675345 *	-0.514 ± 9.68E-02	8.84E-07	-0.542 ± 9.48E-02	2.14E-07

**Table 3: Replicated GWAS associations in admixed populations.** Estimated main genetic effect sizes and p-values of replicated ( $p_{adj} < 0.05$ ) NHGRI associations in admixed individuals (MX: Mexican, PR: Puerto Rican; FVC: forced vital capacity). ADGLM p-values are smaller than LR+PC p-values at 11 of 12 SNPs (starred).

Population	Phenotype	SNP	Effect size ( $\sigma_g^2$ )	P-value
Mexican	Hives	rs705415	1.64 ± 0.33	1.57E-08
	FEV1	rs77775114	-0.5 ± 0.09	3.85E-08
Puerto Rican	FEV1	rs10880087	0.6 ± 0.10	1.39E-08
	FEV1	rs77934571	0.64 ± 0.11	1.18E-08
	FEV1	rs1391405	-0.48 ± 0.09	5.12E-09
	FEV1	rs1640275	0.41 ± 0.07	1.53E-09
	FEV1	rs2567607	-0.52 ± 0.11	2.20E-08
	FVC	rs117344403	36.85 ± 1.02	2.71E-17
	Height	rs117344403	38.87 ± 0.83	3.41E-25
Mexican and Puerto Rican	FEV1	rs55837614	-0.522 ± 0.10	1.16E-09
	FEV1	rs4742581	-0.276 ± 0.10	2.19E-08
	FVC	rs74773809	0.7372 ± 0.10	3.30E-09
African-American	Asthma	rs59843451	0.75 ± 0.14	2.09E-07
	Height	rs12908497	-0.58 ± 0.12	1.84E-08
	Height	rs17712224	-0.34 ± 0.06	2.27E-08
	Height	rs58053243	-0.36 ± 0.07	9.76E-09
	Tanning	rs3877136	-65.51 ± 1.42	1.23E-08

**Table 4: Variance quantitative trait loci in admixed individuals.** Estimated genetic variance effects, standard errors, and p-values of vQTLs ( $p_{adj} < 5e-8$ , adjusted for genomic control) in admixed populations (FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity). Tests of continuous phenotypes, which are below the dotted line within each population sub-table, were run on quantile-normalized phenotypes.