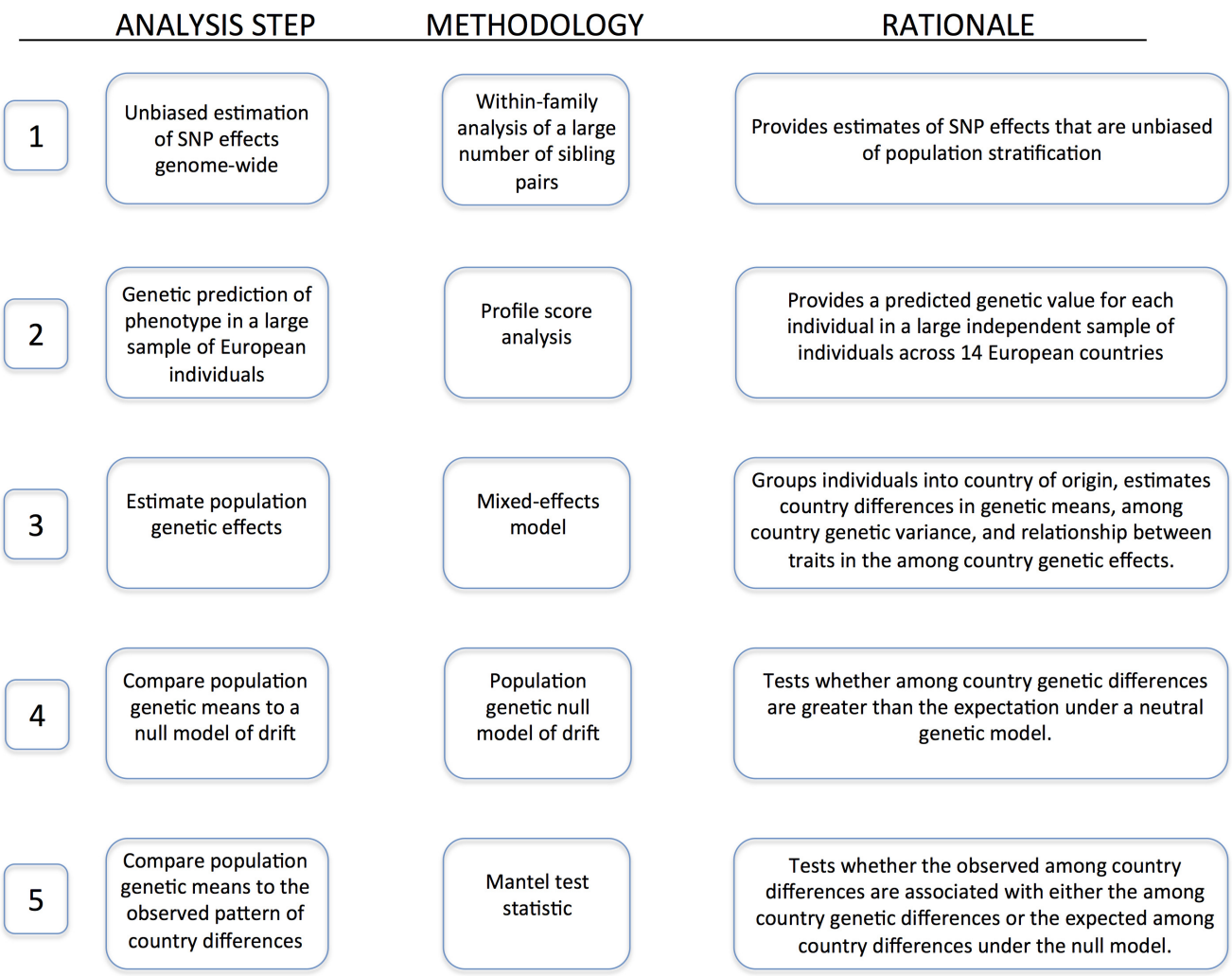
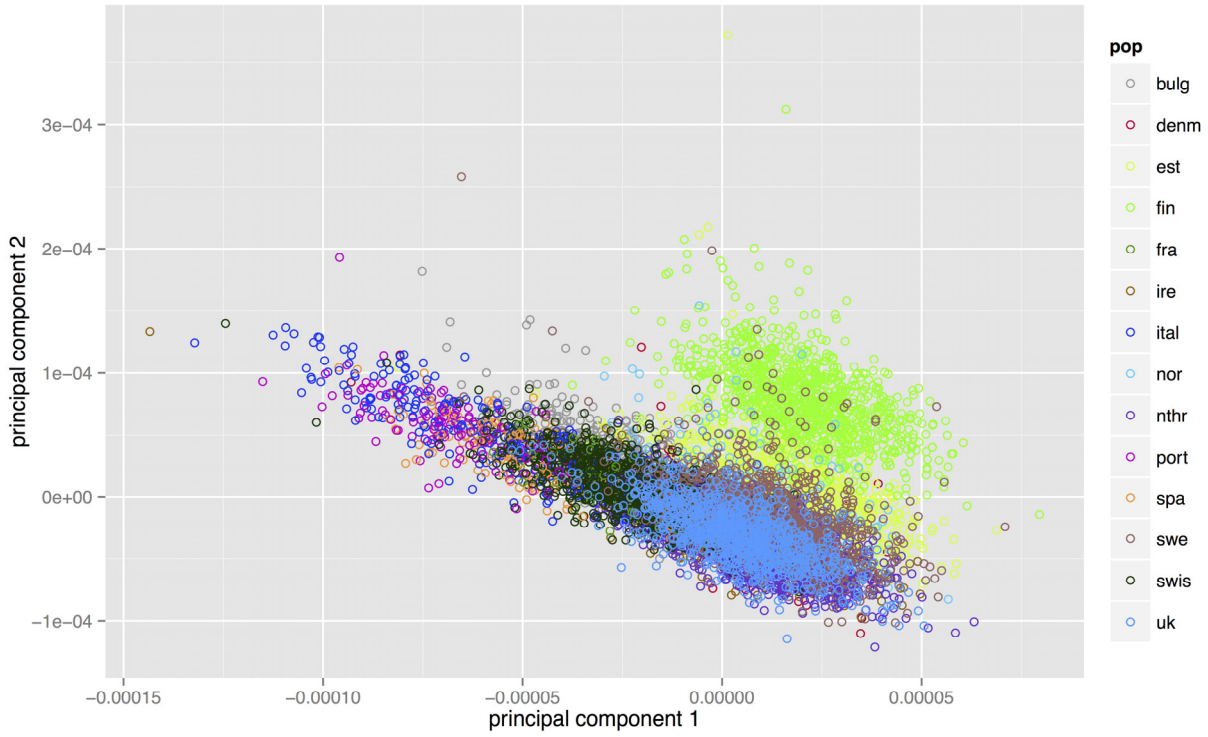


Supplementary Figure 1

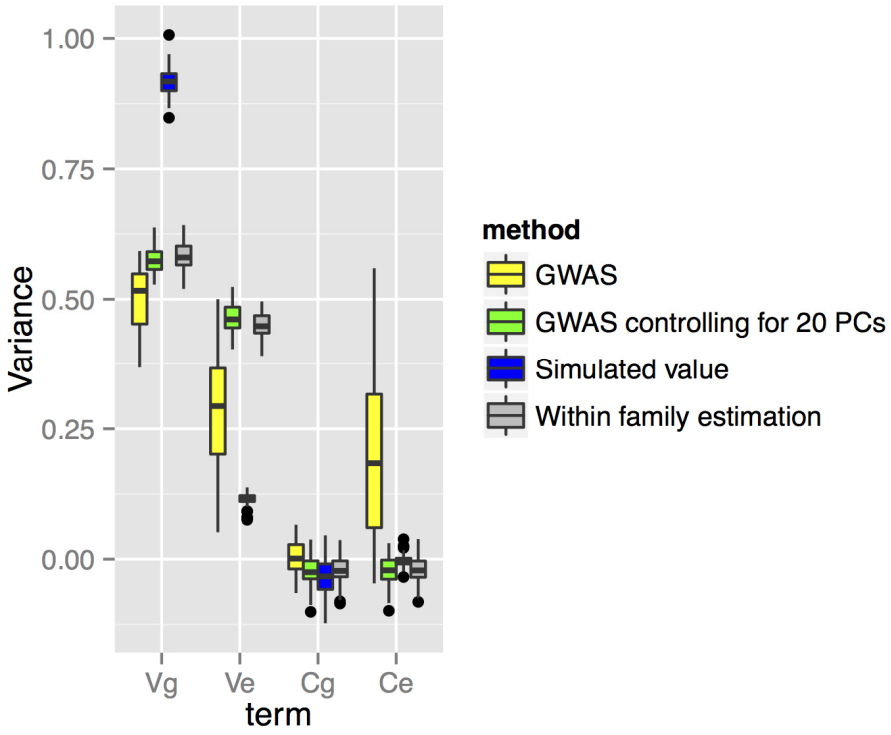
Overview of the study design.



Supplementary Figure 2

Projection of the prediction samples onto the first two HapMap principal components.

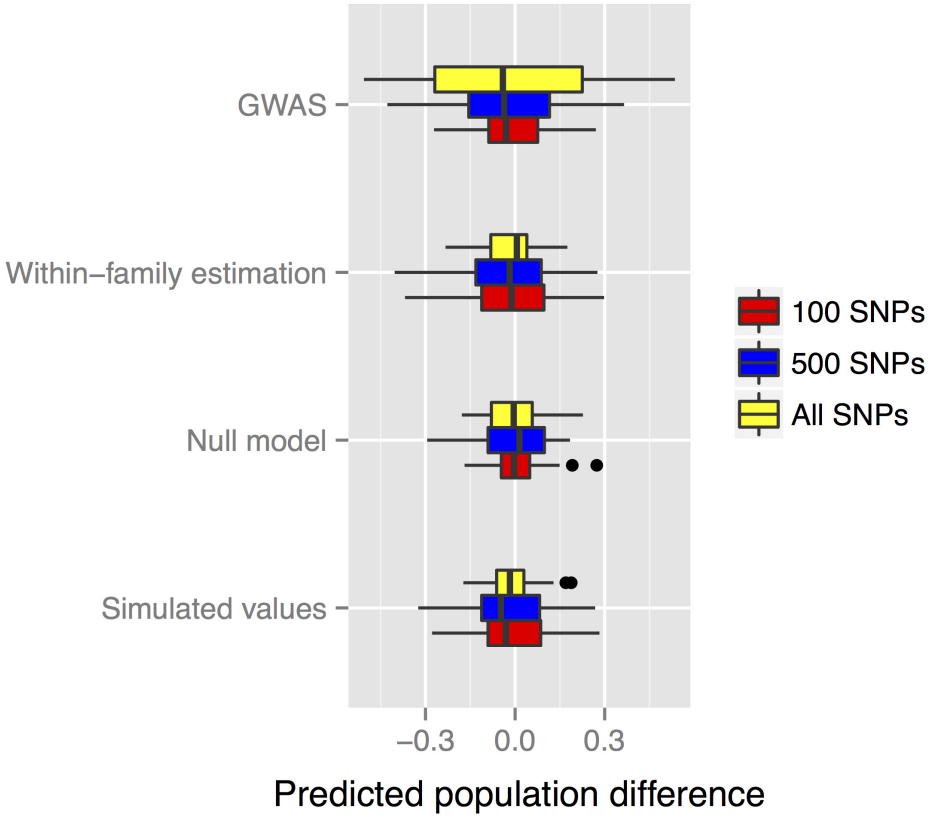
For the non-ascertained set of independent (LD r^2 < 0.1), common, HapMap 3 loci that we used for prediction, we projected our prediction samples onto the first two principal components from HapMap.



Supplementary Figure 3

Simulation study comparing population and within-family association testing.

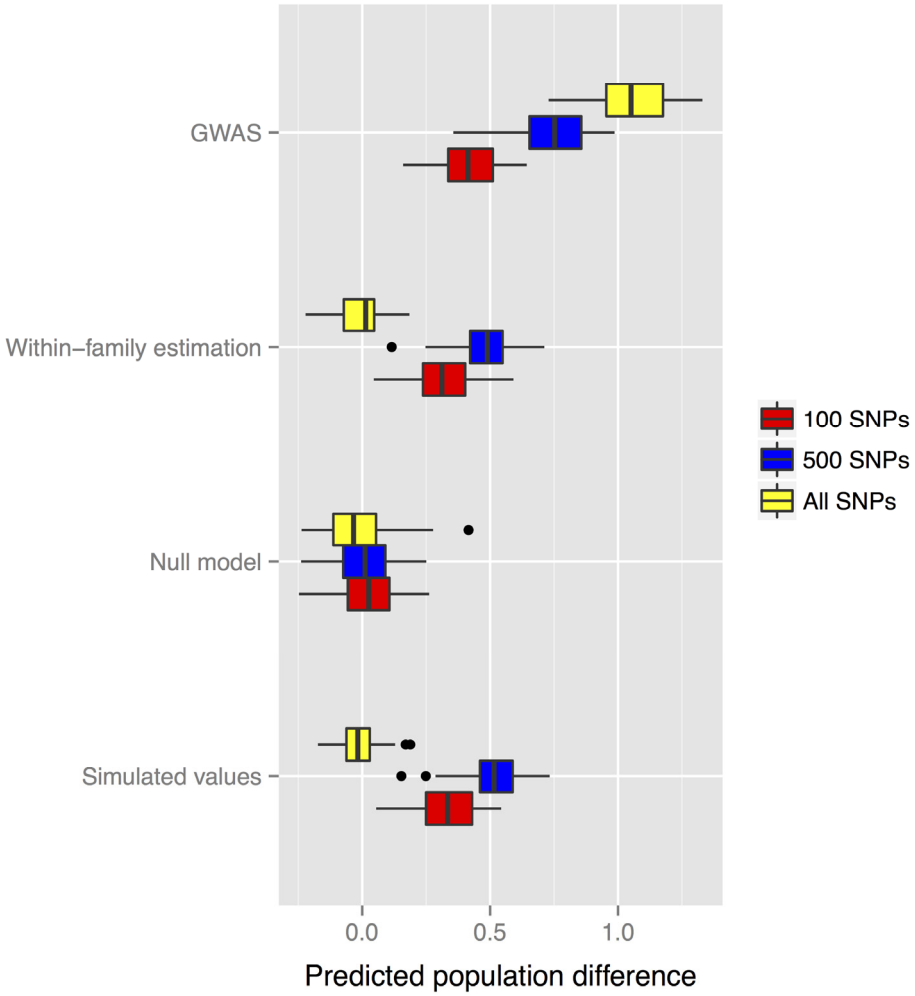
Using real genotype data, causal variants were allocated to 5,000 independent loci at random across the genome, with effects sampled from a normal distribution. Effect sizes were estimated in a genome-wide association study without controlling for population stratification (GWAS; yellow), in a GWAS that controlled for the first 20 principal components (green) and in a sibling pair within-family design (blue). The effect sizes were then used to predict the phenotype in an independent set of sibling pair data, and a recently derived approach was used to test for population stratification bias in the effect size estimates (Online Methods and ref. 33). Variance attributable to a Cg or a Ce term is indicative of population stratification bias, which can be observed in the GWAS scenario that did not appropriately control for population stratification.



Supplementary Figure 4

Simulation study comparing population and within-family association.

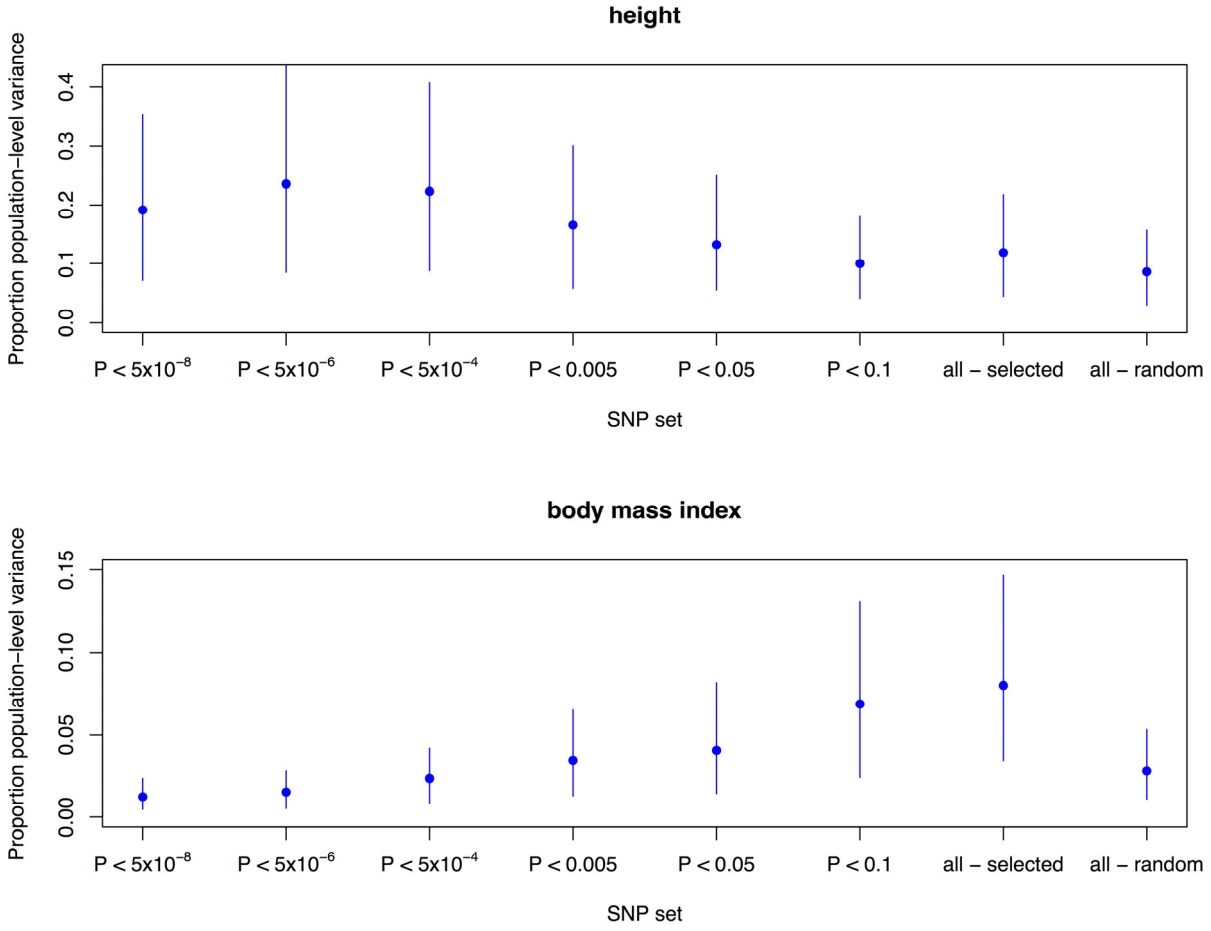
Using real genotype data, causal variants were allocated to 5,000 independent loci at random across the genome, with their effects sampled from a normal distribution. Fifty simulation replicates were conducted. Effect sizes were estimated in a genome-wide association study without controlling for population stratification (GWAS) and in a sibling pair within-family design. The effect sizes from the top 100 independent SNPs identified by the GWAS, the top 500 independent SNPs identified by the GWAS and genome-wide independent loci were then used for prediction in an independent sample. Individuals from the prediction sample were projected onto the first principal component of the discovery sample, and then two groups were selected based on the upper and lower quartiles of the distribution of the projected principal component. The mean difference in the predictor is shown when the predictor was created using effect sizes estimated in a GWAS that did not control for population stratification, using effect sizes estimated in a within-family design, using the true simulated effect sizes and when the within-family effect sizes were randomly allocated to loci under our null model. A predictor created from within-family estimates of effect size yields similar estimates to the true simulated values and the null model, and in no simulation were our predictions significantly different from our null model. This result was irrespective of whether there was ascertainment of loci from a discovery GWAS containing population stratification bias, as when we selected the top 100 or top 500 loci from the GWAS, re-estimated the effects in a within-family analysis and created a predictor we found no evidence of differentiation from our null model. A predictor created from GWAS estimates of effect size that are biased by population stratification yields variable estimates of the prediction difference between the 2 groups, and in 10 of the 50 simulations our predictions differed significantly from our null model.



Supplementary Figure 5

Simulation study comparing population and within-family association testing showing potential ascertainment bias.

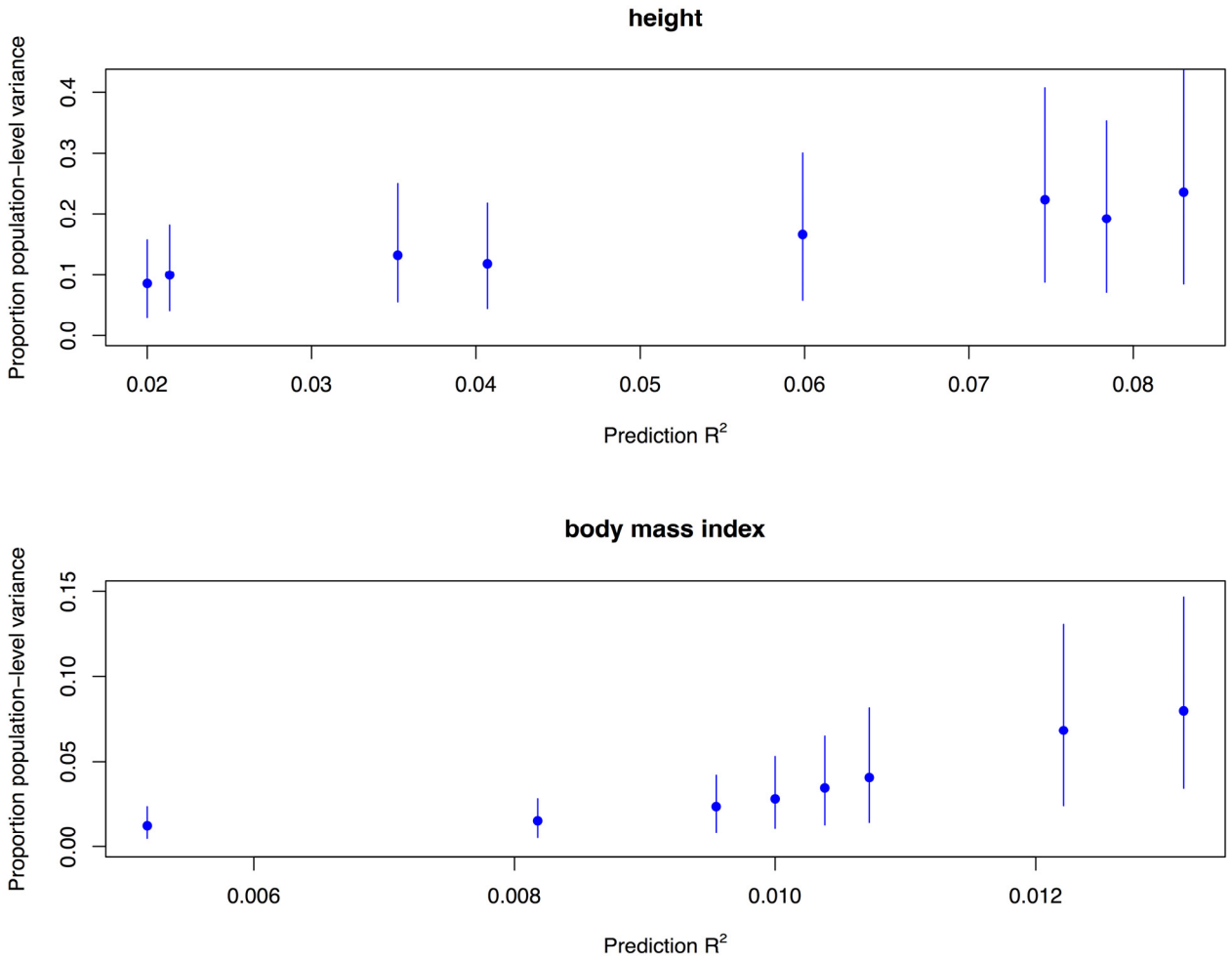
Using real genotype data, causal variants were allocated to 5,000 independent loci at random across the genome, with their effects sampled from a normal distribution. Fifty simulation replicates were conducted. Genotype-environment correlation was induced through a phenotypic mean difference of 0.5 s.d. along the first principal component of the discovery sample. Effect sizes were estimated in a genome-wide association study without controlling for population stratification (GWAS) and in a sibling pair within-family design. The effect sizes from the top 100 independent SNPs identified by the GWAS, the top 500 independent SNPs identified by the GWAS and genome-wide independent loci were then used for prediction in an independent sample. Individuals from the prediction sample were projected onto the first principal component of the discovery sample, and two groups were then selected based on the upper and lower quartiles of the distribution of the projected principal component. The mean difference in the predictor is shown when the predictor was created using effect sizes estimated in a GWAS that did not control for population stratification, using effect sizes estimated in a within-family design, using the true simulated effect sizes and when the within-family effect sizes were randomly allocated to loci under our null model. Ascertainment bias is evident here, where if the top 100 or 500 loci from a GWAS containing population stratification were selected to create a predictor then, irrespective of whether biased SNP effect estimates from the GWAS or unbiased SNP effect estimates from the within-family analysis were used, a significant deviation from the null model is observed. This is because the loci selected from the GWAS were those where the genotype-environment correlation was the strongest. In this simulation scenario, there is no selection on the phenotype and, thus, when there is no ascertainment of loci (when genome-wide SNPs are used to create the predictor), no prediction differences from the true values or the null model were evident.



Supplementary Figure 6

Proportion of population-level variance in a genetic predictor comprised of different sets of SNPs for height and BMI.

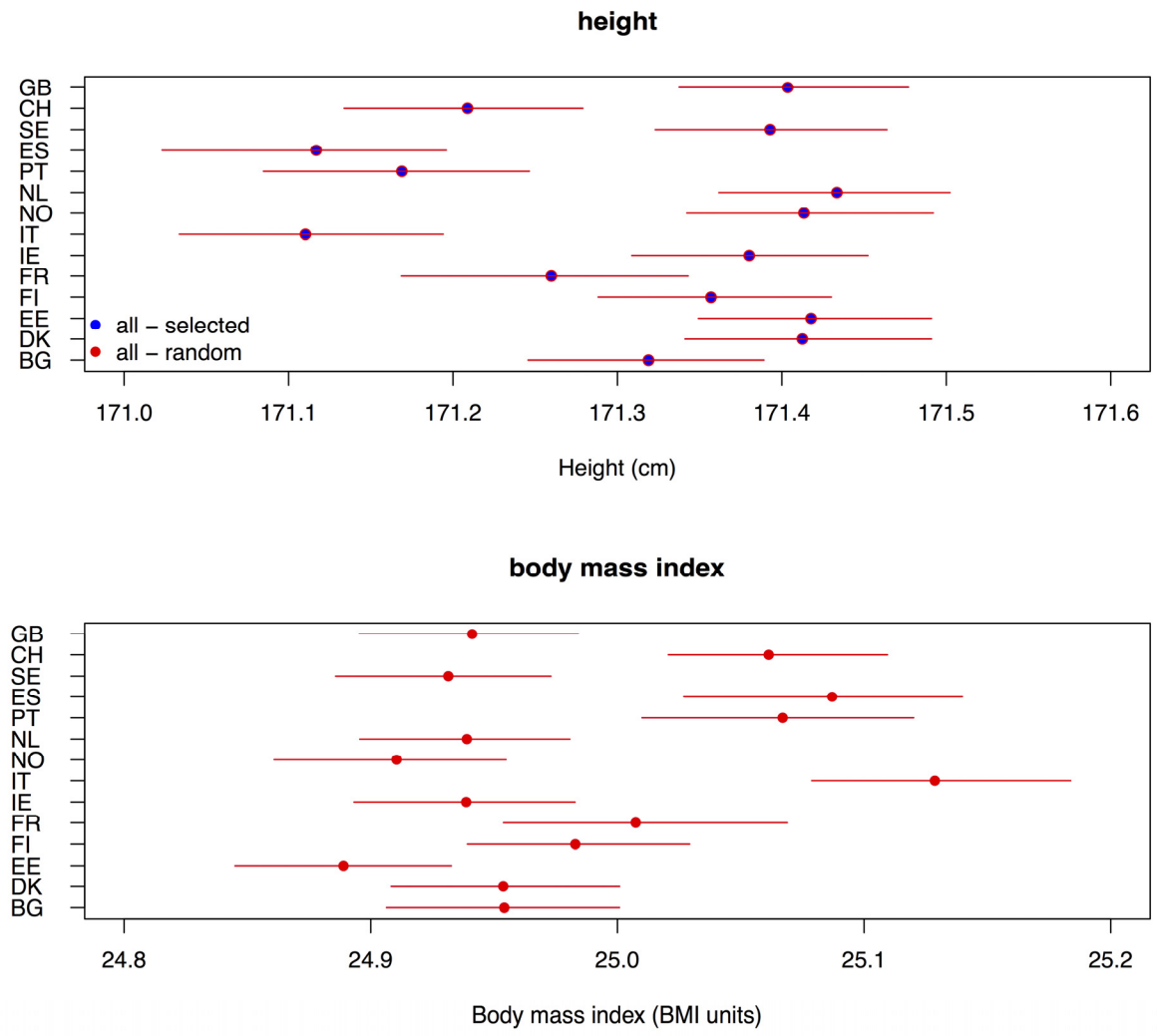
Genetic predictors were created from independent (pairwise LD correlation < 0.1, >1 Mb apart), common HapMap 3 loci selected at different significance thresholds ($P < 5 \times 10^{-8}$, $P < 5 \times 10^{-6}$, $P < 5 \times 10^{-4}$, $P < 0.005$, $P < 0.05$, $P < 0.1$) from large-scale meta-analyses. All refers to genome-wide independent (pairwise LD correlation < 0.1, >1 Mb apart), common HapMap 3 loci that were either selected preferentially on the basis of their within-family association with either trait or selected at random.



Supplementary Figure 7

Increased prediction accuracy results in increased population-level variance.

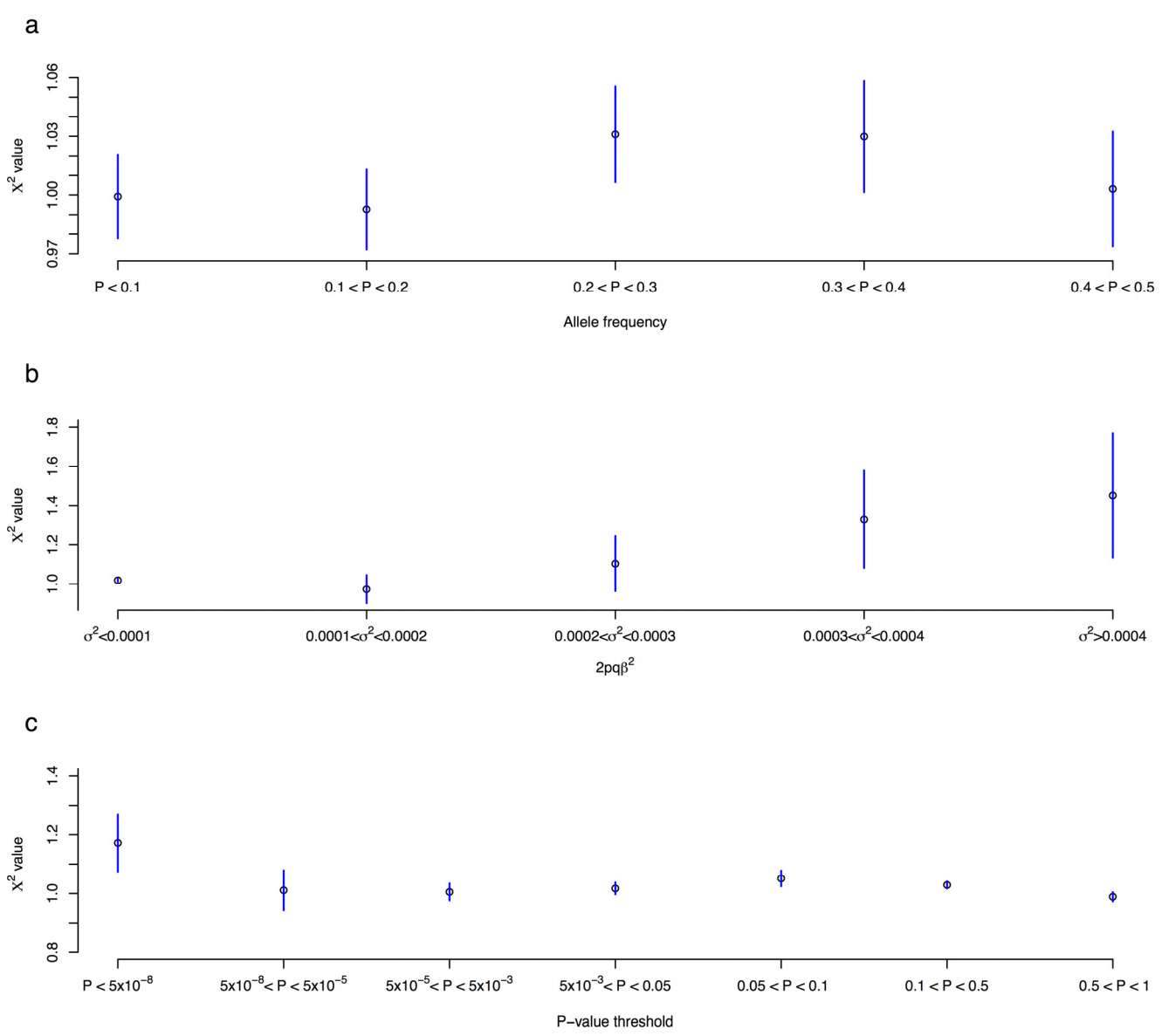
We selected SNPs from large-scale meta-analyses, re-estimated their effects in a within-family design that is unbiased of population stratification and then assessed prediction accuracy in an independent sample for (**a**) height and (**b**) body mass index. We find that the population-level variance is better captured when a greater amount of phenotypic variation is explained by the predictor.



Supplementary Figure 8

Genome-wide pattern of population genetic differentiation for height and body mass index.

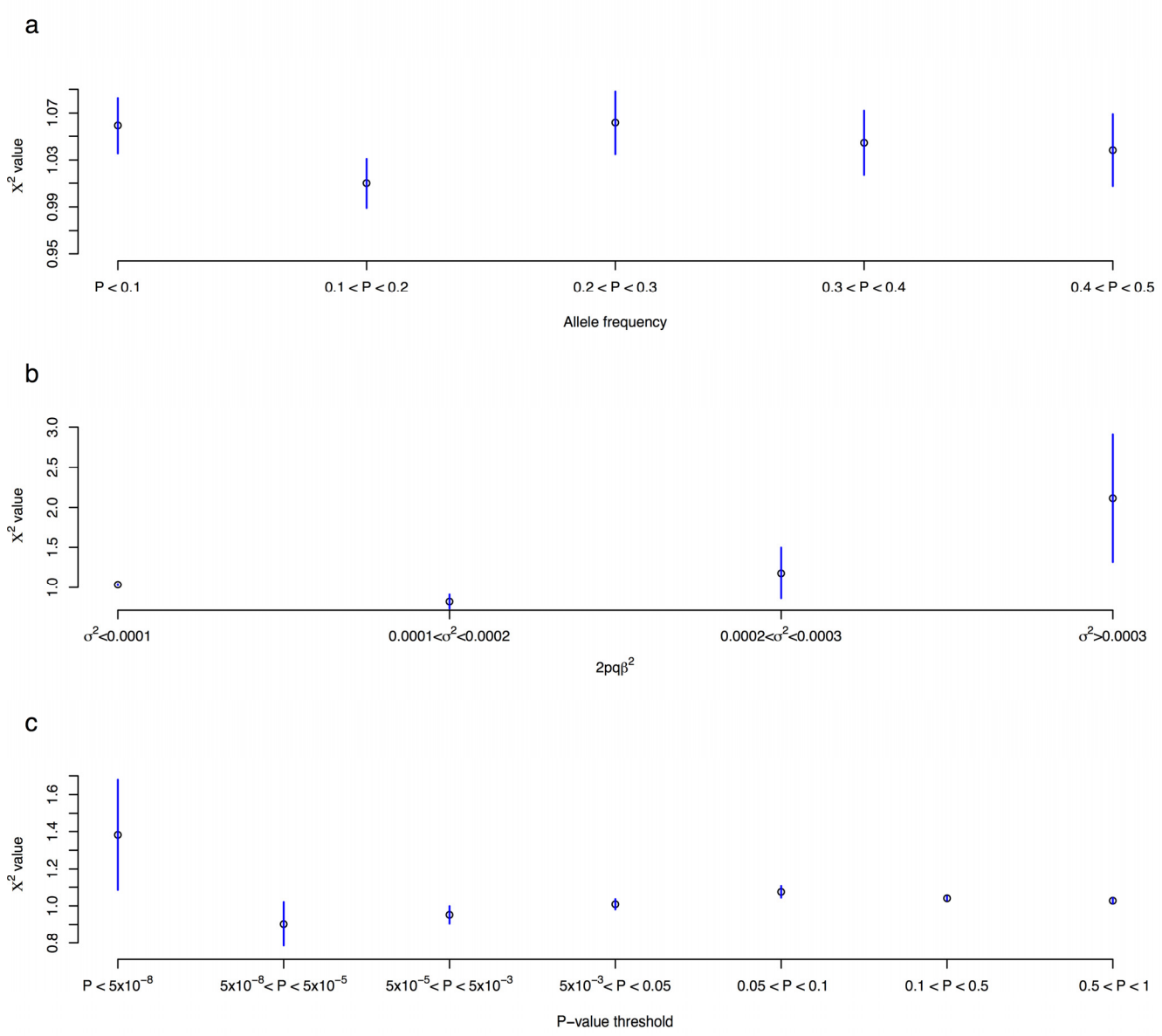
。



Supplementary Figure 9

The characteristics of a SNP and its contribution to the genome-wide pattern of population differentiation for height.

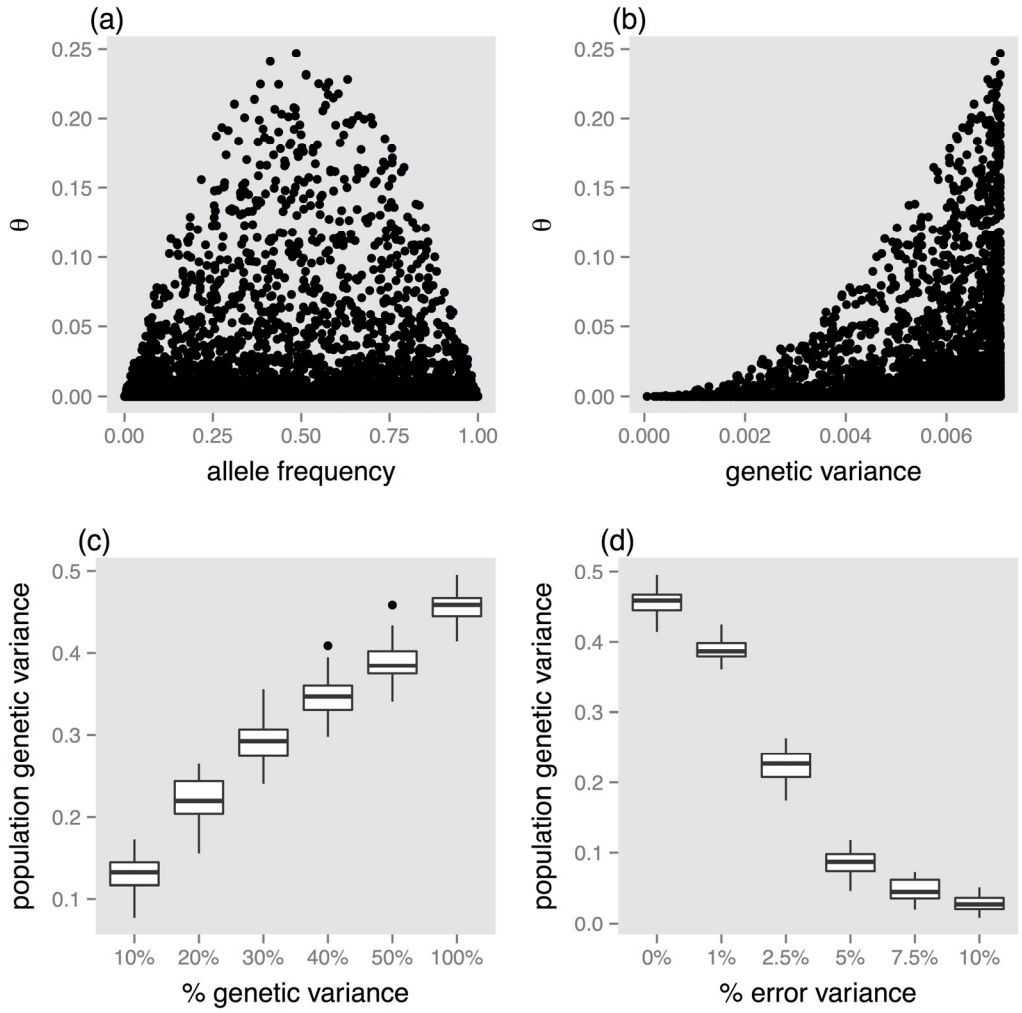
Relationship between the contribution of a SNP to the genome-wide pattern of population differentiation (χ^2 value) and (**a**) allele frequency, (**b**) phenotypic variance explained and (**c**) meta-analysis *P* value.



Supplementary Figure 10

The characteristics of a SNP and its contribution to the genome-wide pattern of population differentiation for body mass index.

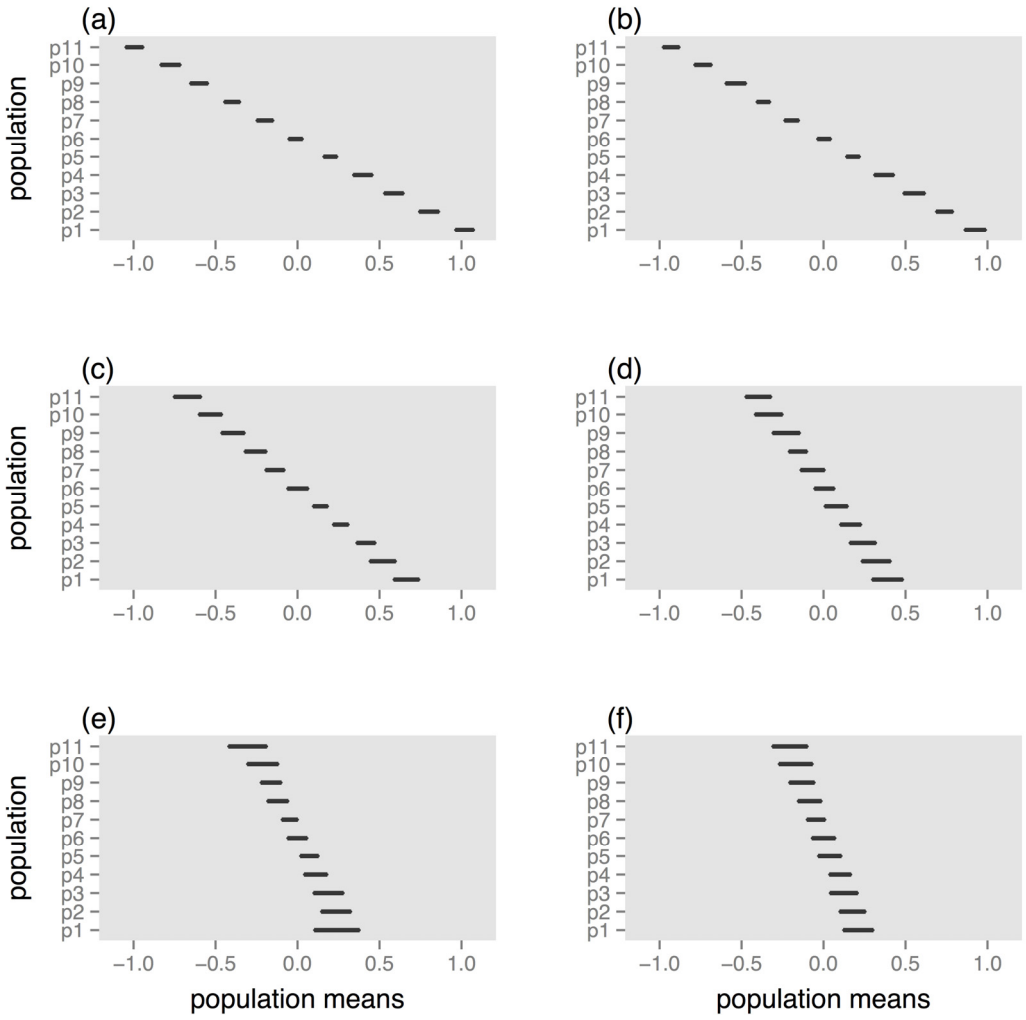
Relationship between the contribution of a SNP to the genome-wide pattern of population differentiation (χ^2 value) and (**a**) allele frequency, (**b**) phenotypic variance explained and (**c**) meta-analysis *P* value.



Supplementary Figure 11

Simulation study of the general approach.

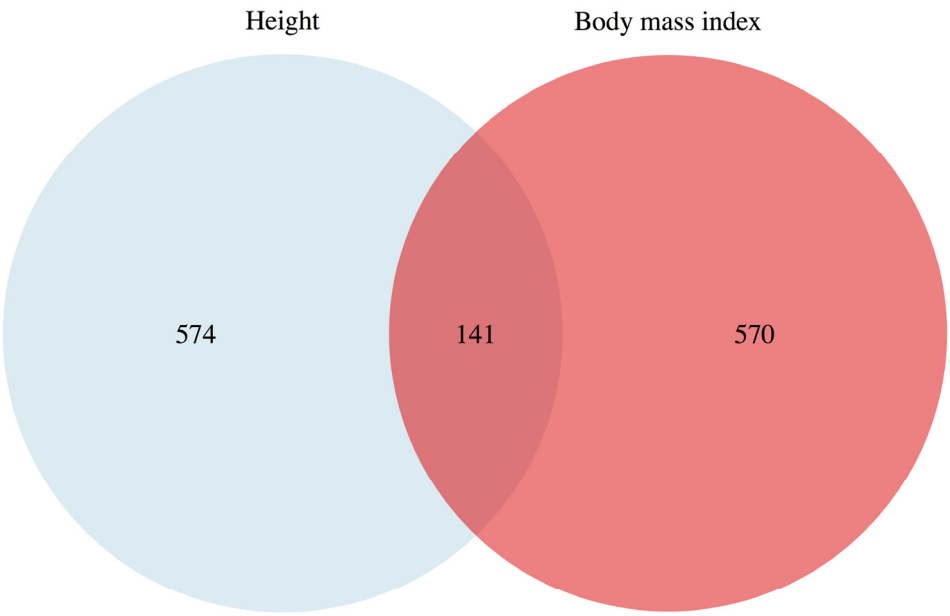
(**a**) Simulated distribution of allele frequency differentiation, *θ*, at 10,000 loci plotted against allele frequency. (**b**) The simulated association between *θ* and the additive genetic variance contributed by each locus, which assumes that the most differentiated loci are those that contribute most to the additive genetic variance. (**c**) Profile scores were calculated from different sets of simulated loci, where each set explained differing amounts of the total variance. Population-level variance is shown for each set of loci, demonstrating that, even when the loci cumulatively explain only 20% of the total variance, 50% of the population-level genetic variance is captured. (**d**) Error variance for each locus was simulated from a normal distribution with variance equal to a percentage of the total variance in profile score. Increasing amounts of error variance were added to the profile score to approximate the effects of adding an increasing number of false positive loci. Population-level variance is shown at different error variances, demonstrating that including a large number of false positives decreases the population-level effects.



Supplementary Figure 12

Simulation study of the error variance induced by the addition of null SNPs.

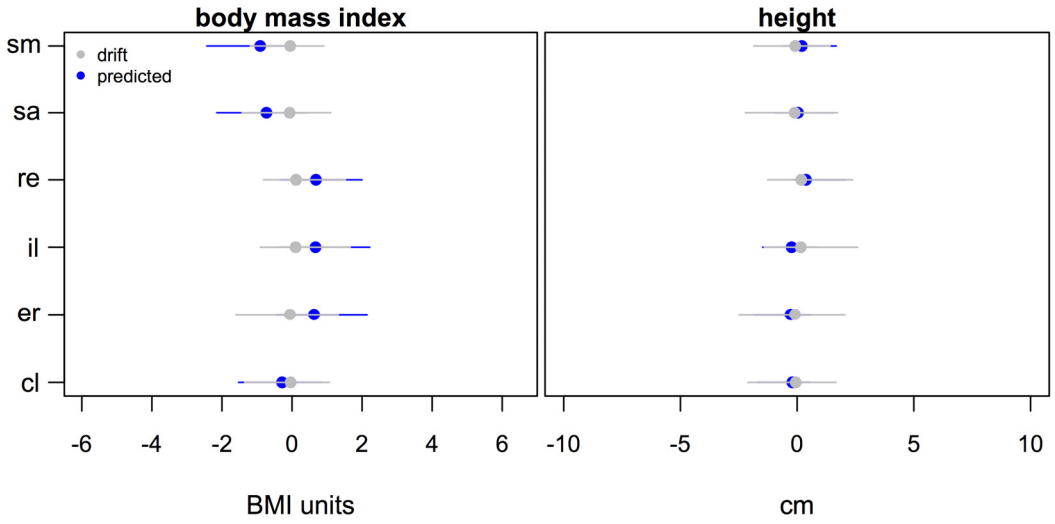
The error variance for each locus was simulated from a normal distribution as a percentage of the total variance in profile score. Increasing amounts of error variance were added to the profile score to approximate the effects of adding an increasing number of false positive loci. The 95% confidence intervals are shown for the population mean profile score of 11 simulated populations across increasing amounts of incorporated error variance: (**a**) 0%, (**b**) 1%, (**c**) 2.5%, (**d**) 5%, (**e**) 7.5% and (**f**) 10%. This pattern reflects the fact that including a large number of false positives decreases the amount of population-level variance estimated.



Supplementary Figure 13

Overlap in the annotation of differentiated loci to genes for height and BMI.

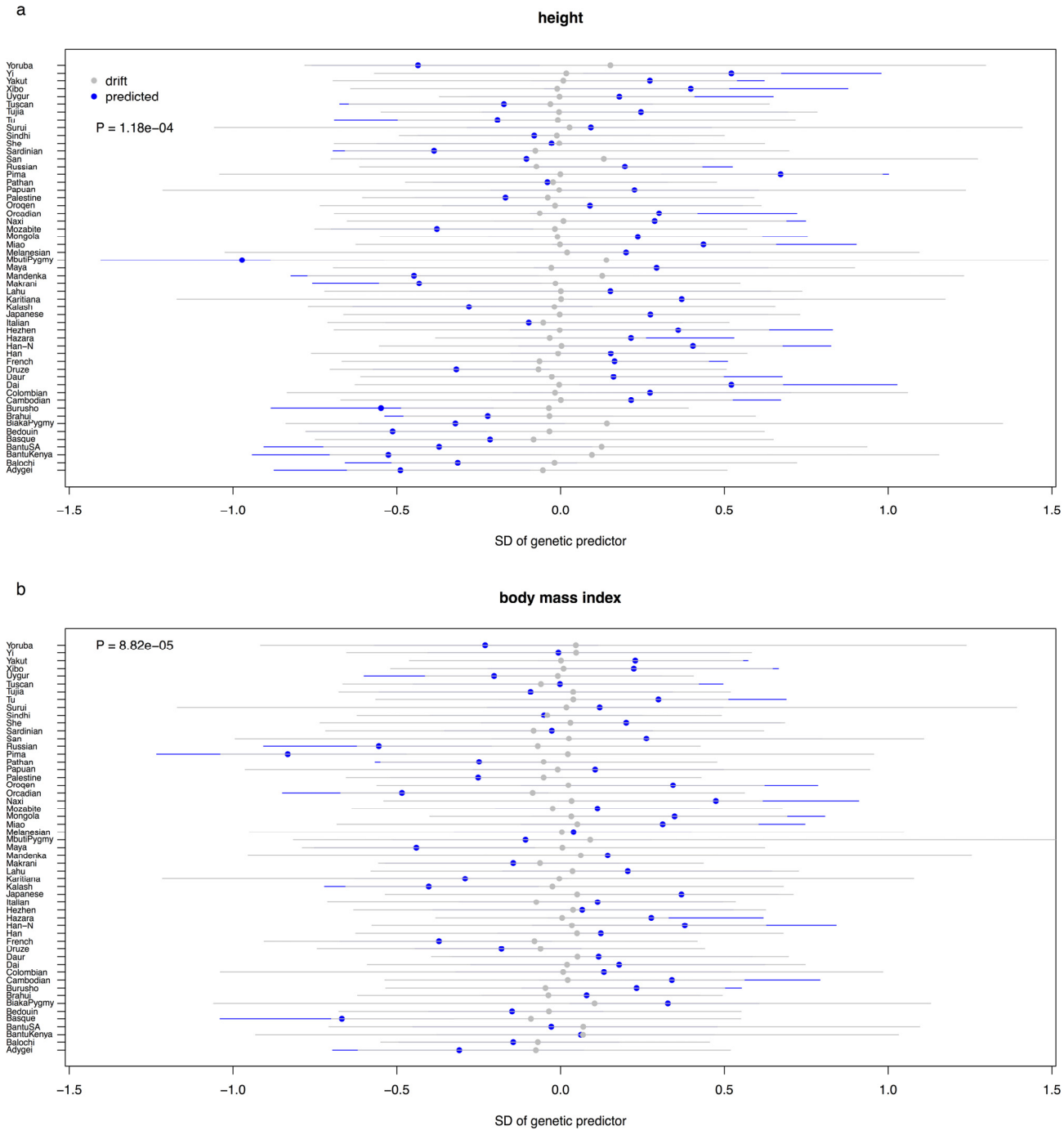
Five hundred SNP loci were selected that are expected to contribute most to the pattern of population genetic differentiation for height and body mass index. These SNP loci were annotated to genes, and the overlap was estimated. The annotation was then repeated by randomly selecting 500 loci from the top 10,000 SNP loci 100 times. The 95% confidence interval of the percentage of overlapping genes across the 100 sampling steps was 8.4–19.4%. In the expected top 500 loci contributing to population genetic variation of each trait, the percentage overlap was 19.8%.



Supplementary Figure 14

Population genetic differentiation across six Italian villages.

Predicted population genetic differentiation for BMI and height on a small scale across six northern Italian villages. There was no significant differentiation from the null model for height (χ^2 = 0.23, P = 0.985) and for BMI (χ^2 = 6.27, P = 0.817) at any set of SNPs, and we present the results across the genome. The villages are Sauris (sa), Resia (re), Illegio (il), Erto (er), Clausetto (cl) and San Martino del Corso (sm) in the Friuli-Venezia Giulia region in northeastern Italy. On this small scale, we find no evidence for population differentiation for either phenotype.



Supplementary Figure 15

Population genetic differentiation in the Human Genetic Diversity Panel for independent genome-wide SNPs ascertained on within-family effect sizes.

(**a**) Height and (**b**) BMI. *P* values give the deviation of the predicted means from the null expectation. The proportion of population-level variance was 17.5% (95% CI = 9.6, 27.9) for height and 13.1% (95% CI = 7.2, 21.3) for BMI. Worldwide, there was no evidence of any population-level genetic correlation (0.007, 95% CI = –0.051, 0.064).

Country	N	Genotyping	Sample source	Reference
EU nations				
Bulgaria	608	Affymetrix 6.0	ISC-Cardiff controls	Lee *et al.* 2012. Nat Genet 44: 247-250
Denmark	458	Illumina 610	SGENE-Copenhagen controls	Lee *et al.* 2012. Nat Genet 44: 247-250
Estonia	1000	Illumina CNV370v1	EGCUT study	Nelis *et al.* PLoS One 5: e5472
Finland	1000	Illumina CNV370v1	STAMPEED: NFBC1966	Sabatti *et al.* 2009. Nat Genet 41(1): 35-46.
France	88	Affymetrix 500K	POPRES study	Nelson *et al.* 2008. Am J Hum Genet 83(3): 347-358
Ireland	860	Affymetrix 6.0	ISC-Trinity College Dublin controls	Lee *et al.* 2012. Nat Genet 44: 247-250
Italy	222	Affymetrix 500K	POPRES study	Nelson *et al.* 2008. Am J Hum Genet 83(3): 347-358
Italian FVG subset	574	Illumina 370CNVv1	FVG Italian study	Esko *et al.* 2013. Eur J Hum Genet 21 (6): 659-665
Netherlands	1000	Ilumina OmniExpress	MinE ALS GWAS	Van Rheenen *et al.* in review
Norway	403	Affymetrix 6.0	SGENE-TOP3 controls	Lee *et al.* 2012. Nat Genet 44: 247-250
Portugal	125	Affymetrix 500K	POPRES study	Nelson *et al.* 2008. Am J Hum Genet 83(3): 347-358
Spain	127	Affymetrix 500K	POPRES study	Nelson *et al.* 2008. Am J Hum Genet 83(3): 347-358
Sweden	1000	Affymetrix 6.0	ISC-SW1 and SW2 controls	Lee *et al.* 2012. Nat Genet 44: 247-250
Switzerland	951	Affymetrix 500K	POPRES study	Nelson *et al.* 2008. Am J Hum Genet 83(3): 347-358
United Kingdom	1000	Affymetrix 6.0	WTCCC2: 1958 British Birth Cohort controls	WTCCC2 2011. Nature 476:214-219
Prediction accuracy	12596	Illumina Omni-2.5	HRS: Health and Retirement Study	http://hrsonline.isr.umich.edu/index.php?p=regcou
Within-family estimation	17500	Various	Sib-pair Consortium data	Hemani *et al.* 2013. Am J Hum Genet 93(5): 865-875

Supplementary Table 1. Genomics data used. Sample sizes (N), genotyping platform, sample source and references.

GO ID	top 500	null	P	Term
height				
GO:0007156	20	3	0.000497444	Plasma membrane adhesion
GO:0007399	20	6	0.008531169	Nervous system development
GO:0051276	11	2	0.016535842	Chromosome organization
GO:0005509	29	13	0.019584876	Calcium ion binding
GO:0042802	15	6	0.058794869	Identical protein binding
GO:0046904	8	2	0.071422272	Calcium ion binding
GO:0005516	8	2	0.071422272	Calcium binding protein
GO:0006886	6	1	0.077809732	Intercellular protein transport
GO:0010008	6	1	0.077809732	Endosome membrane
GO:0030234	22	12	0.096849544	Enzyme regulator activity
GO:0005829	44	29	0.100511365	Cytosol
GO:0005506	7	2	0.112796195	Iron iron binding
GO:0000122	13	6	0.1161557	RNA polymerase II promoter
GO:0005667	5	1	0.130803461	Transciptor factor complex
GO:0020037	5	1	0.130803461	Iron heme binding
GO:0031982	5	1	0.130803461	Vescicle
GO:0044325	5	1	0.130803461	Ion channel binding
GO:0043167	111	86	0.131720435	Ion binding
GO:0005575	236	194	0.133485716	Cellular component
GO:0005623	207	169	0.136637148	Cell
body mass index				
GO:0043025	16	5	0.017161404	Neuronal cell body
GO:0016020	62	38	0.018222939	Cellular membrane
GO:0007399	14	4	0.01946807	Nervous system development
GO:0030425	12	3	0.021646301	Dendrite cells
GO:0004930	11	3	0.034598998	G-protein receptor activity
GO:0007267	26	13	0.036345473	Cell to cell signalling
GO:0007186	14	5	0.039363886	G-protein receptor signalling
GO:0016323	7	1	0.040353413	Basolateral plasma membrane
GO:0030424	8	2	0.06303948	Axon cellular component
GO:0048011	8	2	0.06303948	Neorotrophin TRK receptor signalling pathway
GO:0051301	8	2	0.06303948	Cell division
GO:0005615	24	13	0.064892639	Extracellular space
GO:0016021	81	60	0.076886846	Membrane component
GO:0007411	15	7	0.081651885	Axon guidance
GO:0043005	9	3	0.08448119	Neuron projection
GO:0016874	7	2	0.101429936	Ligase activity
GO:0007268	17	9	0.103459888	Synaptic transmission
GO:0005525	5	1	0.120309564	GTP binding
GO:0007612	5	1	0.120309564	Learning
GO:0030165	5	1	0.120309564	PDZ domain binding

Supplementary Table 2. Top GO annotations for loci expected to contribute most to the pattern of population genetic differentiation for height and body mass index.

Supplementary Note

1. Estimating population-level genetic variance

One locus, two populations. Following[32,47,51–53], if two populations are descended from a common ancestor, we can describe variance in the frequency of an allele *k* across populations as:

$$\sigma_{p,k}^2 = (\frac{1}{n_r}) \sum_{r=1\ldots n_r} \left(p_{k,r} - \bar{p}_k\right)^2 \qquad [1.1]$$

where $p_{k,r}$ is the frequency of allele k within each population r, $\bar{p}_k$ is the allele frequency across the entire sample and n_r is the number of populations.

The ratio of the variance of allele frequencies among populations relative to the total variance in allele frequency of *k* can be described as:

$$\theta_k = \frac{\sigma_{p,k}^2}{\bar{p}_k(1-\bar{p}_k)} \qquad [1.2]$$

which is equivalent to the classical estimator of F_{ST} describing allelic differentiation at a locus across populations[52].

Consider a quantitative trait that is influenced by a single allele *k,* the genetic variance contributed the locus is:

$$\sigma_{g,k}^2 = 2\bar{p}_k(1 - \bar{p}_k)a_k^2 \qquad [1.3]$$

where a_k is the effect size on an arbitrary scale of allele *k*, and $\bar{p}_k$ is the frequency of the allele *k* within the sample.

Assuming that a_k is equal across the two populations, the among population variance is:

$$\sigma_{r,k}^2 = 2\bar{p}_k(1 - \bar{p}_k)a_k^2\theta_k \qquad [1.4]$$

where θ_k is the allelic differentiation across the two populations estimated in Eq. 1.2.

The ratio of these two components at this locus is:

$$Q = \frac{\sigma_{r,k}^2}{\sigma_{g,k}^2} = \theta \qquad [1.5]$$

which describes the proportion of genetic variance attributable to population-level effects for allele *k*.

Thus for a causal variant, the proportion of additive genetic variance attributable to population-level effects is equal to estimating the amount of allelic

Nature Genetics: doi:10.1038/ng.3401

differentiation, θ, at that locus[51]. Eq. 1.5 can be written another way if we consider the mean genetic value of a population for an allele:

$$g_{k,r} = 2p_{k,r}a_k \quad [1.6]$$

the mean genetic value across populations is then:

$$g_k = 2\bar{p}_k a_k \quad [1.7]$$

and the genetic variance across populations is then:

$$\sigma^2_{r,k} = (\tfrac{1}{n_r}) \sum \left(2p_{k,r}a_k - 2\bar{p}_k a_k\right)^2 \quad [1.8]$$

with the ratio of the variance across populations to the additive genetic variance within populations as:

$$Q = \frac{\sigma^2_{r,k}}{4\bar{p}_k(1-\bar{p}_k)a^2_k} = \theta \quad [1.9]$$

Multiple loci and multiple populations. If we now consider a quantitative trait that is influenced by a set of n_k alleles where x_k is an indicator function for the allele k (x= 0, 1, 2 copies for diploid individuals) of individual i. Considering only additive effects, the additive value for a trait of individual i is a sum of the locus specific effects across the n_k alleles:

$$g_i = \sum_k x_{k,i}a_k \quad [1.10]$$

Additive variance at the population level is then:

$$\sigma^2_{\bar{g}} = (\tfrac{1}{n_r}) \sum_{X \in r} (\bar{g}_r - \bar{g})(\bar{g}_r - \bar{g})^T, \quad [1.11]$$

where $\bar{g}_r$ is the mean genetic value within each population r, and $\bar{g}$ is the mean genetic value over all individuals in the sample. The ratio of the population-level variance to the total variance is then:

$$Q = \frac{\sigma^2_{\bar{g}r}}{\sigma^2_{\bar{g}}} \quad [1.12]$$

which describes the proportion of genetic variance attributable to population-level effects across loci.

The quantity Q in Eq. 1.12 is equivalent to estimating the cumulative effects of θ across the loci that influence a quantitative trait. However, as there are multiple loci it now contains two components:

$$Q = \frac{\sum_k \sigma^2_{p,k}a^2_k}{\sum_k r(\bar{p}_k(1-\bar{p}_k)a^2_k)} + \frac{\sum_j \sum_k a_k a_j Cov(p_k, p_j)}{\sum_k (\bar{p}_k(1-\bar{p}_k)a^2_k)} \quad [1.13]$$

Nature Genetics: doi:10.1038/ng.3401

where the first component is the average θ across all loci, and the second component is the covariance between alleles *k* and *j* in their directional deviation multiplied by the product of their effects a_k and a_j, summed across loci, and then divided by their additive contribution to the trait variance.

Under drift the second component is expected to be zero, as neutral loci will not all consistently covary in the same direction as the effects of the loci, with variation around the expectation depending on the number of loci.

Under selection, the second component will be >0 because selection creates a match between effect size direction and frequency differences across the trait-associated SNPs. Therefore for a trait under differential selection across populations, trait-associated loci are expected to have higher values of θ than neutral loci, but they will also covary in a consistent manner across loci, creating consistent mean differences in profile score across populations.

2. Using estimated SNP effect sizes from GWAS

Population-level variance in a genetic predictor. GWAS results provide estimates of the additive value at a locus, which can then be used to create a genetic predictor $\hat{g}_i = \sum_k x_{k,i}\hat{\beta}_{k,m}$, where $\hat{\beta}_k$ is the standardized regression coefficient of the SNP allele *k*.

Considering a single trait, we can create a genetic predictor for individuals across multiple independent data sets from different regions, and then the genetic predictor can be decomposed into components:

$$\hat{g}_i = \mu + v_r + e_i, \qquad [2.1]$$

where μ is the mean of the genetic predictor (global mean genetic value), v_r is the population-level additive genetic effects with value $\bar{g}_r$ for each population *r* and variance $\sigma^2_{v_r}$, and e_i is the residual variance representing the average individual-level effects within populations. This approach is only valid if the regression coefficients used to create the genetic predictor are unbiased of population stratification.

Eq. 2.1 can be estimated in a Bayesian MCMC mixed-effects model framework, to model the amount of population-level variance in the genetic predictor $\sigma^2_{v_r}$, and to gain model estimates of the population-level genetic values $\bar{g}_{r,m}$. An MCMC approach provides multiple posterior estimates of the parameters, which is important as it allows uncertainty in the parameter estimates to be carried through to all of the later analyses described below. This is important especially if there is an imbalance in sample size among countries, because the greater uncertainty in the mean estimates of smaller sample sizes will be taken into account. If profile scores are created using different sets of SNPs, then comparisons among sets can be made by standardizing the genetic predictor to a z-score prior to analysis provides an estimate of the deviation of each individual,

Nature Genetics: doi:10.1038/ng.3401

and thus each population, in SD. We can then calculate the proportion of variance attributable to among population effects as $\frac{\sigma^2_{\hat{v}_r}}{\sigma^2_{\hat{v}_r}+\sigma^2_{\hat{e}}}$.

When multiple phenotypes have been recorded, the population-level variance and covariance can be decomposed as:

$$\widehat{\boldsymbol{g}} = \boldsymbol{\mu} + \boldsymbol{v} + \boldsymbol{e}, \qquad [2.2]$$

where $\widehat{\boldsymbol{g}}$ is a matrix of genetic predictors of individual *i* for traits *m*, $\boldsymbol{\mu}$ is a vector of trait mean genetic values, $\boldsymbol{v}$ is the population-level additive genetic value for *m* traits, and $\boldsymbol{e}$ is the residual variance representing individual-level effects for *m* traits. Again, Eq. 2.2 can be estimated in a multivariate Bayesian MCMC mixed-effects framework, to model the population-level variance for all traits *m*, the population-level covariance (co-differentiation), and the residual covariance, which represents the within-population covariance in genetic predictors (within population genetic correlation).

Estimation and prediction bias. The regression coefficients used to create the genetic predictor must be unbiased of population stratification; otherwise a quantification of population genetic differentiation may be biased. We quantify this bias using a hypothetical design with two populations.

Consider a stratified discovery sample consisting of equal proportions of individuals from two diverged sub-populations, and a causal SNP x_k, which has an effect on phenotype of β_k. The allele frequency of x_k is $\bar{p}_k = 0.5\left(p_{1_k} + p_{2_k}\right)$, where p_{1_k} and p_{2_k} are the frequencies of SNP k in populations 1 and 2 respectively. We define the difference in allele frequency among populations as $\delta_k = p_{1_k} - p_{2_k}$ and $F_{ST_k} = \frac{0.5\left(p_{1_k}+p_{2_k}\right)^2}{\bar{p}_k(1-\bar{p}_k)}$.

We denote the population means as $\bar{y}_1$ and $\bar{y}_2$. We define the difference in the population means as $\bar{y}_1 - \bar{y}_2 = 2\beta\left(p_{1_k} - p_{2_k}\right) + \Delta$, where the first term is the contribution from the variant to the population difference and the second term, Δ, is a non-genetic effect. If we ignore population stratification in the discovery GWAS:

$$\hat{\beta}_k = \frac{\sigma(y,x_k)}{\sigma^2_{x_k}},$$
$$\sigma^2_{x_k} = E(x_k^2) - E(x_k)^2 = 2\bar{p}_k(1-\bar{p}_k)\left(1 - 0.5F_{ST_k}\right),$$

dropping subscript *k* for convenience. Since,

$$E(x^2) = \frac{1}{2}\left(\sigma^2_{x_1} + E(x_1)^2 + \sigma^2_{x_2} + E(x_2)^2\right),$$

then,

$$\sigma(x,y) = E(xy) - E(x)E(y),$$
$$= \beta\left(p_1(1-p_1) + p_2(1-p_2)\right) + \frac{1}{2}(\bar{y}_1 - \bar{y}_2)(p_1 - p_2),$$

Nature Genetics: doi:10.1038/ng.3401

$$= \beta\big(p_1(1-p_1) + p_2(1-p_2)\big) + \frac{1}{2}(p_1-p_2)(2b(p_1-p_2) + \Delta),$$
$$= b(2\bar{p}(1-\bar{p}))\left(1+\frac{1}{2}F_{ST}\right) + \frac{1}{2}\Delta(p_1-p_2),$$

and hence,

$$\hat{\beta}_\delta = \beta + \frac{\frac{1}{2}\Delta(p_1-p_2)}{(2\bar{p}(1-\bar{p}))\left(1+\frac{1}{2}F_{ST}\right)} \qquad [2.3]$$

with bias: $c = \frac{\frac{1}{2}\Delta(p_1-p_2)}{(2\bar{p}(1-\bar{p}))\left(1+\frac{1}{2}F_{ST}\right)}$.

Across the genome the expectation of *c* for all SNPs is zero because the direction of allele frequency differentiation, will not match the direction of phenotypic differentiation across all SNPs. However, stratification bias in the regression coefficient estimates from the discovery GWAS, may create a match between effect size direction and frequency differences across SNPs identified as being genome-wide significant. This is because loci are more likely to be identified if the stratification in allele frequency differences matches both the difference in phenotypic mean among populations and the direction of effect size.

We can describe the non-central χ^2 distribution with mean and variance of a test statistics (T) of association conditional on all parameters (including the allele frequency difference between the two populations) as:

$$E(T|\delta) = 2\bar{p}(1-\bar{p})N\beta_\delta^2 \qquad [2.4]$$

$$\sigma^2(T|\delta) = 2 + 4\bar{p}(1-\bar{p})N\beta_\delta^2 \qquad [2.5]$$

where $\delta = p_1 - p_2$. The expectation over $\hat{\beta}_\delta{}^2$ over δ,

$$E_\delta(\beta_\delta^2) = E\left(\beta + \frac{1}{4}\Delta(\bar{p}(1-\bar{p}))(p_1-p_2)\right)^2,$$
$$= \beta^2 + (\bar{y}_1 - \bar{y}_2)^2 F_{ST}/8\bar{p}(1-\bar{p}) \qquad [2.6]$$

Hence, the expectation of the non-centrality-parameter (NCP) of detection, with the expectation taken over δ is,

$$E_\delta(NCP) = 2\bar{p}(1-\bar{p})N\left(\beta^2 + \frac{(\bar{y}_1 - \bar{y}_2)^2 F_{ST}}{8\bar{p}(1-\bar{p})}\right),$$
$$= Nq^2 + {}^1\!/_4\, N(\bar{y}_1 - \bar{y}_2)^2 F_{ST} \qquad [2.7]$$

with q^2 the proportion of variance explained by the variant. The variance of the NCP is:

$$\sigma_\delta^2(NCP) = 4\bar{p}(1-\bar{p})N\left(\beta^2 + \frac{(\bar{y}_1 - \bar{y}_2)^2 F_{ST}}{8\bar{p}(1-\bar{p})}\right),$$
$$= 2Nq^2 + {}^1\!/_2\, N(\bar{y}_1 - \bar{y}_2)^2 F_{ST} \qquad [2.8]$$

Scaled by N, the mean and variance of the NCP are q^2 + ¼$N(\bar{y}_1 - \bar{y}_2)^2 F_{ST}$ and $2q^2$ + ½$N(\bar{y}_1 - \bar{y}_2)^2 F_{ST}$, respectively. If we consider a polygenic phenotype that is stratified across two countries, with $\bar{y}_1 - \bar{y}_2$= 1 SD, effect sizes of ~ 0.014 SD or less, and the variance explained (q^2) of each SNP typically less than 0.1%, then even if $F_{ST} \sim 0.01$ across Europe, the mean and variance of the NCP will be mostly driven by stratification. This means that if population 1 has a greater mean height, then SNPs where the height increasing allele has a higher frequency in population 1 are far more likely to be detected than those where the height increasing allele is not differentiated between the populations. Conversely, SNPs where the height decreasing allele has a lower frequency in population 1 are also far more likely to be detected than non-differentiated SNPs. Therefore, the top trait-associated loci will have higher values of θ than neutral loci, and they will covary in a consistent manner across loci, creating consistent mean differences in profile score across populations, that is due to stratification bias in the GWAS (ascertainment bias), rather than selection unless population stratification is accounted for in the discovery GWAS.

If we consider U to be a dummy variable for sub-population, with U=1 for population 1 and 0 for population 2, then F_{ST} at each locus can be interpreted as the proportion of variance in x_k explained by U, so $F_{ST_k} = R^2(x_k, U)$, with R^2 the squared correlation between x_k and U. Then essentially we can consider the problem as the effect of a SNP is biased due to the confounding effect of U. The confounding effect will be $(\bar{y}_1 - \bar{y}_2) r \sqrt{\frac{\sigma_U^2}{\sigma_{x_k}^2}}$, with r the correlation of x_k and U which is analogous to Eq. 2.3 above.

If there is stratification, a correlation will be created between the SNP regression coefficients, $\hat{\beta}$, and r, which will create a biased genetic predictor with mean differences in the predictor between the sub-populations. Thus, when creating a genetic predictor as the weighted sum of the top SNP effects for two populations:

$$\hat{g}_i = \sum_k x_{k,i}\hat{\beta}_k + \sum_k \frac{\frac{1}{2}\Delta\left(p_{1_k} - p_{2_k}\right)}{2p_k(1-p_k)(1+\frac{1}{2}F_{ST_k})} \qquad [2.9]$$

where the bias at each of the ascertained SNPs sum in a directional manner because SNPs where $\bar{y}_1 - \bar{y}_2$ and r are in the same direction (those that are differentiated in the same direction as the phenotype) get upwardly biased; and those where $\bar{y}_1 - \bar{y}_2$ and r are in opposing directions get downwardly biased.

In GWAS, the effect of each SNP is tested whilst controlling for the leading principal components, with the aim of removing the confounding effects of U. While this may control for a great deal of population stratification, there is no certainty that all stratification has been removed from the SNP regression

Nature Genetics: doi:10.1038/ng.3401

coefficients of large-scale meta-analyses. Therefore, an alternative approach is to use a within-family analysis to estimate the SNP regression coefficients, avoiding population stratification biases, and then to confirm results using a predictor at non-ascertained genome-wide loci to ensure that conclusions gained are not just a result of ascertainment biases.

3. Within-family analysis association analysis

There are numerous approaches to test association in family-based designs, where the aim is to estimate regression coefficients that are unbiased of population stratification. One approach is to partition association effects into orthogonal between and within-family components in which the former reflects population structure and the latter is only significant in the presence of linkage disequilibrium[65]. For sibling pairs, the model is:

$$y_i = \mu + \beta_{b,k} b_k + \beta_{w,k} w_k + e_i, \qquad [3.1]$$

where y_i is the phenotype measured on individual i, and b_k and w_k are orthogonal between- and within-family components for a given SNP k, and e is the residual error. b_k is calculated as $b_k = \frac{\sum x_{k,f}}{n_f}$, where n_f is the number of members in a family and $\sum x_{k,f}$ is the sum of the SNP values of all the members of a family. w_k is calculated as $w_k = x_k - b_k$ and is thus the deviation from the family mean of each individual. The regression coefficient $\hat{\beta}_{w,k}$ is thus a direct estimate of the additive genetic value of a marker that is unbiased of stratification bias.

A genetic predictor can then be made using the regression coefficients $\hat{\beta}_{w,k}$ as:

$$\hat{g}_i = \sum_k x_{k,i} \hat{\beta}_{w,k} \qquad [3.2]$$

The effect of LD can be minimized or eliminated by using a set of independent SNPs in linkage equilibrium.

4. Comparison of the population-level estimates to a null drift model

Testing selection from drift genome-wide. To compare the population-level estimates to their expectation under random genetic drift, a quantitative genetic framework for studying population differentiation can be used[47].

For SNP k, genotyped in population 1 and 2, the difference in allele frequency estimates approximately follows a normal distribution:

$$\delta_k = p_{1_k} - p_{2_k} \sim N\left(0, \bar{p}_k(1 - \bar{p}_k)\left(2F_{ST} + \frac{1}{2N}\right)\right) = N(0, \sigma_\delta^2) \qquad [4.1]$$

Weighting the SNP by its regression coefficients $\hat{\beta}_{w,k}$ produces a genetic predictor for each SNP, g_k, where the difference in mean between populations also approximately follows a normal distribution.

$$\delta_k = \bar{g}_{1_k} - \bar{g}_{2_k} \sim N\left(0, \left(\frac{1}{n_r}\right)\sum_{X\in r}(\bar{g}_1 - \bar{g})(\bar{g}_2 - \bar{g})^T\right) = N(0, \sigma_g^2) \qquad [4.2]$$

Under drift, the expectation of the mean is zero across all SNPs genome-wide because drift would not create a consistent alignment of the direction of allele frequency differentiation and effect size across all loci. If there is selection, then the mean will no longer be zero and this is what we wish to test across multiple populations.

Following[47], the matrix of population-level effects estimated in Eq. 2.2 can be denoted as $\boldsymbol{A}^r = (\bar{g}_{r,m})_r$. If we consider a population consisting of a set r of n_r local populations, with each population consisting of n_i individuals, which are derived from a common ancestral population, then under drift $\boldsymbol{A}^r$ is expected to be distributed as multivariate normal:

$$\mathbf{A}^r \sim MVN(\boldsymbol{\mu} \otimes \mathbf{I}, 2\mathbf{G}^A \otimes \boldsymbol{\theta}^r), \qquad [4.3]$$

where $\boldsymbol{\mu}$ is the mean additive genotype in the ancestral population which is determined by the allele frequencies in the ancestral population, $\mathbf{I}$ is a unit vector relating observations to populations, $\mathbf{G}^A$ is the ancestral variance-covariance matrix of the traits in question, $\boldsymbol{\theta}^r$ is the matrix of population-level coancestry, and $\otimes$ is the Kroenecker product. Relatedness can be defined as the probability that randomly chosen alleles from a given locus of individuals i and i' are identical by descent, with average coancestry between any two populations X and Y denoted by $\theta^r_{XY} = (1/n_X n_Y)\sum_{i\in X, i\in Y}\theta_{ii}$, which make the off-diagonal elements of the matrix $\boldsymbol{\theta}^r$, and average within-population relatedness $\theta^r_X = \theta^P_{XX}$ on the diagonal. Therefore, genetic values across local populations $\boldsymbol{A}^r$ are assumed to be multivariate normal and dependent upon the degree of relatedness to other populations, the ancestral additive variance-covariance of the traits, and the ancestral trait means.

While we know nothing of the ancestral genetic values, we can estimate the expected population-level values under drift and then compare our predicted values to these using the framework of Eq. 4.3. This requires a number of steps:

(i) Randomize the regression coefficients $\hat{\beta}_{w,k}$ across SNPs. By keeping the effect sizes consistent, but attributing those effects across SNPs at random, our profile scores reflect the action of drift.

(ii) Transform each set of profile scores to a z-score and using them in Eq. 2.2 to provide 1000 estimates of the population genetic variance, and population means under drift denoted as $\boldsymbol{A}^D$. These values are displayed in the figures as the neutral model estimates.

Nature Genetics: doi:10.1038/ng.3401

(iii) Calculating the sample covariance matrix of these 1000 estimates, which provides an estimate of the expected population-level covariance in phenotype under drift, denoted as $\mathbf{\Sigma}$.

(iv) Using a Mahalanobis distance statistic to provide a relative measure of the deviation of our predicted population-level means from their multivariate theoretical expectation under drift[47]. This provides the χ^2 test statistic, used to compare our predicted estimates to their drift expectation.

For (iv), the Mahalanobis distance of $\boldsymbol{A}^r$ from their multivariate expected value under drift $\boldsymbol{A}^D$ is $D = \sqrt{(\boldsymbol{A}^r - \boldsymbol{A}^D)^T \mathbf{\Sigma}^{-1}(\boldsymbol{A}^r - \boldsymbol{A}^D)}$. There is a single mean estimate of $\mathbf{\Sigma}$, but parameter uncertainty is accounted for by calculating D 1000 times, using the 1000 MCMC estimates of $\boldsymbol{A}^r$ and to the 1000 estimates of $\boldsymbol{A}^D$. D^2 follows a χ^2 distribution with degrees of freedom equal to the number of traits and populations, which thus provides a test of whether the predicted population-level genetic values deviate from a drift expectation. This is a common approach used to detect multivariate outliers. Because both the drift profile scores and the trait profile score are transformed to a z-score this comparison is on the same SD scale.

Although the assumption of normality for drift allele frequencies may be violated if there is substantial drift (which is not the case within Europe), we can appeal to the central limit theorem, as across many loci the estimated genetic values may still be assumed to follow a multivariate normal. Additionally, although our predictions will be imperfect predictions of the true genetic values this does not invalidate our null model. Our null model describes the expected allele frequency change under neutrality, which is independent of any ascertainment in the GWAS and independent of LD among SNPs because we are using within-family regression coefficients at independent SNP loci. Thus, neutrally evolving SNP loci will be well described by a comparison of their predictions to those made by randomly allocating effect sizes across SNPs, irrespective of any violations in our model assumptions.

Testing population genetic differentiation of a single SNP. Having determined whether population differentiation of a genetic predictor differs to the pattern expected under drift, we can then extend this framework to identify individual loci that contribute to the population genetic effects.

In Eq 4.2, $\bar{g}_{1_k} - \bar{g}_{2_k} \sim N(0, \sigma_\delta^2)$ and by standardizing the predictor of each SNP k to a variance of 1, $\bar{g}_{1_k} - \bar{g}_{2_k} \sim N(0, 1^2)$. From Eq 2.2, we have estimated a population effect for each population, v_r, representing estimates of the differences among populations in a predictor for height and BMI. The cross-product of the standardized predictor of each SNP, $\hat{g}_k$ and a vector of standardized population effects, v_r, gives a standard normal density:

$$\delta_k = \frac{\sum_N \hat{g}_k v_r}{\sqrt{k}} \qquad [4.4]$$

where ${\delta_k}^2$ follows a χ_1^2. Therefore, we gain a statistic of the contribution of each SNP to the identified pattern of population genetic. If the genome-wide pattern of population mean differences are driven by few highly differentiated SNPs then these loci will have a larger χ_1^2 than expected, alternatively if there are a large number of loci of small differentiation that sum up to create the genome-wide pattern than no single locus will significantly differ from expectation.

5. Comparing predicted genetic differentiation to observed phenotypic differentiation

To examine the contribution of genetic effects to the observed phenotypic pattern, a distance matrix can be calculated between each pair of populations with respect to their average recorded phenotype. This observed phenotypic distance matrix $\mathbf{H}^{\boldsymbol{O}}$ is calculated as

$$\mathbf{H}^{\boldsymbol{O}} = \sum_{m=1}^{h_m} \left(y_{r_n,m} - y_{r_{n+1},m}\right)^2 \qquad [5.1]$$

where r is the vector of observed mean phenotypes for n of r populations for trait m. h_m is the number of phenotypes.

A distance matrix $\mathbf{H}^{\boldsymbol{P}}$ for the profile scores of the quantitative traits can also be defined as

$$\mathbf{H}^{\boldsymbol{P}} = \sum_{m=1}^{h_m} \left(\bar{g}_{r_n,m} - \bar{g}_{r_n+1,m}\right)^2 \qquad [5.2]$$

where $\bar{g}_{r_n,m}$ and $\bar{g}_{r_n,m}$ are the predicted population means for n of r populations for trait m. h_m is the number of phenotypes.

The association between $\mathbf{H}^{\boldsymbol{O}}$ to $\mathbf{H}^{\boldsymbol{P}}$ can be determined using a Mantel test statistic[66]:

$$\mathrm{M}(\mathbf{H}^O, \mathbf{H}^P) = \sum_{i=1}^{n_r} \sum_{j=1}^{i} \mathrm{H}_{ij}^O \mathrm{H}_{ij}^P \qquad [5.3]$$

where n_r denotes the number of populations.

The distribution of this test statistic $\mathrm{M}(\mathbf{H}^O, \mathbf{H}^P)$ can be compared to that expected under random genetic drift, by calculating to $\mathbf{H}^{\boldsymbol{P}}$ using the 1000 profile scores created from the randomly selected set of SNPs. When these 1000 drift derived $\mathbf{H}^{\boldsymbol{P}}$ matrices are compared to the observed $\mathbf{H}^{\boldsymbol{O}}$, it provides an expected distribution of the test statistic under drift, denoted $\mathbf{H}^{\mathbf{P,R}}$.

We can then examine the probability that the randomized product moment $\mathrm{M}(\mathbf{H}^O, \mathbf{H}^{P,R})$ is less than the observed statistic $\mathrm{M}(\mathbf{H}^O, \mathbf{H}^P)$

$$H := \mathrm{P}\{\mathrm{M}(\mathbf{H}^O, \mathbf{H}^P) > \mathrm{M}(\mathbf{H}^O, \mathbf{H}^{P,R})\} \qquad [5.4]$$

Nature Genetics: doi:10.1038/ng.3401

Parameter uncertainty is accounted for because H is calculated for each recorded MCMC chain t, with the fraction of cases where $\mathrm{M}(\mathbf{H}^{O}, \mathbf{H}_{t}^{P}) > \mathrm{M}\left(\mathbf{H}^{O}, \mathbf{H}_{t}^{P,R}\right)$ determining the interpretation of the test statistic. A value of H close to one implies that the observed distribution of population means is more similar to the predicted means than would be expected at random, i.e. under random genetic drift. Again, standardizing the observed data, the drift profile scores, and the trait profile scores to a z-score enables $\mathbf{H}^{O}$ to $\mathbf{H}^{P}$ to be compared on the same SD scale.

6. Transforming population means onto the observed scale.

The profile scores, $\hat{g}$ created for each individual i for each trait m, can be approximately transformed back to the observed scale by:

$$w_{i,m} = SD(m)cov(z_y, z_{\hat{g}})z_{\hat{g}} \qquad [6.1]$$

where $\mathrm{SD}(m)$ is the standard deviation of the trait m in the discovery GWAS sample, which is the true SD of the observed phenotype within the population; and $cov(z_y, z_{\hat{g}})$ is the covariance within an independent prediction sample between a z-score of phenotype $y_{i,m}$ and a z-score of the trait profile score $\hat{g}_{i,m}$. Although there will be estimation error, we proceed under the assumption that $cov(z_y, z_{\hat{g}})$ approximately describes the amount of phenotypic variation explained by the profile score within an independent population. To transform back to the observed scale the profile score is thus multiplied by a proxy of the amount of phenotypic variation explained and then multiplied by the SD of the observed phenotype within the population. While z-score comparisons are preferred for the analysis, because they do not rely on transformation and allow comparisons to occur on the same scale, the $w_{i,m}$ values can be used for graphical representation of the predictions relative to the true observed phenotypic differences across nations.

7. Approximations of the parameters in simulation study.

If individuals, and thus genotypes, are sampled randomly across randomly mating populations then the sampling distributions of $\bar{p}_k$, θ_k and a_k can be approximated. Recall from above that the variation among populations at a locus is:

$$\sigma_{r,k}^2 = 2\bar{p}_k(1 - \bar{p}_k)a_k^2\theta_k \qquad [7.1]$$

Values of p_k across multiple loci can be sampled from a uniform beta distribution with shape parameter of 1, which provides a representative distribution of the frequencies of ascertained SNPs in a sample. For the additive genetic effect at a locus, we expect a relationship between the effect size and the allele frequency[55] such that the effect size for each causal variant can be sampled from a normal distribution as:

$$a_k \sim N(0, \sqrt{\frac{h^2}{2p_k(1-p_k)n_k}}) \quad [7.2]$$

where h^2 is the heritability, p_k is the allele frequency, and n_k is the number of causal variants. This approximation give larger effect sizes for loci with lower minor allele frequency, which is supported by theory and empirical observation[15, 67].

θ_k is likely to vary across loci and we can assume greater differentiation at intermediate allele frequencies and sample θ_k as:

$$\theta_j = p_k(1 - p_k)Beta(\alpha, \beta) \quad [7.3]$$

where a Beta distribution is used to sample θ_k as a continuous probability distribution with interval 0 and 1. The beta distribution is parameterized by shape parameters α and β, which we set to 0.05 and 1 giving a uniform distribution with small mean and small probability of high θ_k. This distribution has been used previously to describe allele frequencies in population genetics [ref]. Multiplying this by $p_k(1 - p_k)$ means that the value of θ_k is associated with the frequency p_k, with intermediate frequencies having higher θ_k values, which is supported by theory and empirical evidence[68, 69].

Differences in allele frequencies across populations can then be sampled as:

$$p_{k,r} = p_k + \theta_k \otimes S_r \quad [7.4]$$

where S_r are mean differences among populations in profile score and $p_{k,r}$ gives the frequency of allele k within each population r. This creates small-scale differences among populations in allele frequency across loci that are consistent in direction, creating a value for the second component of Eq. 1.13 that is >0. From these sampled means allelic data can be sampled as:

$$x_{k,i,r} \sim Bin(p_{k,r}, 2) \quad [7.5]$$

where a Binomial distribution is used to sample $x_{k,i,r}$ which is an indicator for the allele k (k= 0, 1, 2 copies for diploid individuals) of individual i in population r, with 2 as the number of trails (2 alleles per diploid individual), with $p_{k,r}$ is the probability of selecting the causal SNP allele at each locus within each population.

The additive value for trait m for individual i within population r is then the sum of the locus specific effects across the n_k loci:

$$\hat{g}_{i,r} = \sum_k x_{k,i,r} a_k \quad [7.6]$$

The values $\hat{g}_{i,r}$ can then be used in Eq. 2.1 to calculate the ratio of the population-level variance to the total variance.

GWAS results provide estimates of the additive value at a locus, but these estimates are made with error. The profile scores of the individuals within the sample thus contains the components:

$$\widehat{g_i} = \mu + v_r + e_i, \qquad [7.7]$$

where $e_i = a_i + \epsilon_i$, with a_i as the true genetic value and an error term ϵ_i which is the estimation error that is summed up across the loci used to create the profile score. This cumulative error is likely to increase as the number of false positive SNPs that are included within the profile score increases. This error can be approximated and included within the sampling as:

$$g_{i,r} = \sum_{k,} x_{k,i,r}\left(a_{k,r,m} + \epsilon_{k,r,m}\right), \quad \epsilon_{k,r,m} = N(0, z) \qquad [7.8]$$

where $\epsilon_{k,r,m}$ is sampled from a normal distribution with standard deviation z.

Eq. 7.1 through Eq. 7.8 approximate the sampling distribution of all parameters considered based on theoretical expectations The total variance of $\hat{g}_i$ can be modified by varying the parameters of h^2; the population-level variance $\sigma^2_{v_r}$ can be modified by varying the shape parameters of the beta distribution used to sample θ_k along with the selection parameter S_r; and the cumulative error variance within the profile score can be modified by altering z in Eq. 7.8. The number of SNPs included in the score can also be modified by ranking SNPs according to their variance explained, and selecting different sets based on the cumulative variance explained, assuming that GWAS SNPs are ascertained based on the variance explained by allele k at locus j as:

$$\sigma^2_{A,k} = 2\bar{p}_k(1 - \bar{p}_k)a_k^2 \qquad [7.9]$$

We performed simulations using Eq. 7.1 through Eq. 7.9. We fixed the proportion of population-level variance across simulations and examined how altering the amount of error variance and the number of loci ascertained influenced the estimates gained.

8. Improvements on previous approaches.

Our approach differs from a previous study[32] in that it estimates population-level variance using individual-level genetic profile scores rather than simply deriving the population-level mean genetic value. This difference is key, because it enables both the amount of population genetic variance and the population means to be estimated relative to the amount of individual-level variance within each population. Additionally: (i) we use estimates of SNP effects that are unbiased of population stratification; (ii) we advocate estimating population–level variance from profile scores calculated from many thousands of SNPs genome-wide rather than only those reaching genome-wide significance thus capturing more of the trait variation and better characterizing the population-level effects; (iii) we adopt a multivariate approach which is appropriate as selection does not act upon single phenotypes independently[47]; and we estimate

the co-differentiation among phenotypes across populations, which can be compared to the individual-level within-population covariance; (iv) we estimate the correlation between the observed pattern of differentiation and the predicted genetic values relative to the expectation under drift; (v) all samples used in this study have European ancestry to minimize loss of LD; (vi) all of our parameters are estimated using an MCMC approach which allows 95% credible intervals to be placed on all estimates and allows uncertainty in the parameter estimates to be carried through to all of our later analyses. In general, this framework overcomes many previously limiting factors when examining population genetic differentiation because: (i) we examine differentiation across multiple genomic regions together, rather than focusing on specific regions, which is appropriate given the likely polygenic nature of genetic variation[26,70]; (ii) we confirm our results in a non-ascertained set of data, thus avoiding the potential for ascertainment biases; and (iii) differences among populations are all assessed relative to each other in a single framework, not just along single linear gradients or between two groups, which is appropriate given that populations radiate away from a common ancestor and each other[47].

Supplementary Note References

65. Abercasis, G.R., Cardon, L.R. & Cookson, W.O.C. A general test of association for quantitative traits in nuclear families. *Am J. Hum Genet* **66**, 279-292 (2000)

66. Karhunen, M., Ovaskainen, O., Herczeg, G. & Merilä, J. Bringing habitat information into statistical tests of local adaptation in quantitative traits: a case study of nine-spined sticklebacks. *Evolution* **68,** 559–68 (2014).

67. Eyre-Walker, A. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc. Natl. Acad. Sci. U. S. A.* **107 Suppl ,** 1752–6 (2010).

68. Elhaik, E. Empirical distributions of F(ST) from large-scale human polymorphism data. *PLoS One* **7,** e49837 (2012).

69. Jakobsson, M., Edge, M. D. & Rosenberg, N. A. The relationship between F(ST) and the frequency of the most frequent allele. *Genetics* **193,** 515–28 (2013).

70. Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20,** R208–15 (2010).