HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2016 November 01.

Published in final edited form as:

Nat Genet. 2015 November ; 47(11): 1357–1362. doi:10.1038/ng.3401.

Population genetic differentiation of height and body mass index across Europe

Matthew R. Robinson[1], **Gibran Hemani**[1], **Carolina Medina-Gomez**[2], **Massimo Mezzavilla**[3,4], **Tonu Esko**[5,6,7,8], **Konstantin Shakhbazov**[1], **Joseph E. Powell**[1,9], **Anna Vinkhuyzen**[1], **Sonja I. Berndt**[10], **Stefan Gustafsson**[11], **Anne E. Justice**[12], **Bratati Kahali**[13], **Adam E. Locke**[14], **Tune H. Pers**[6,7,8,15], **Sailaja Vedantam**[6,7], **Andrew R. Wood**[16], **Wouter van Rheenen**[17], **Ole A. Andreassen**[18], **Paolo Gasparini**[3,4], **Andres Metspalu**[5], **Leonard H. van den Berg**[17], **Jan H. Veldink**[17], **Fernando Rivadeneira**[2], **Thomas M. Werge**[19,20,21], **Goncalo R. Abecasis**[14], **Dorret I. Boomsma**[22,23,24], **Daniel I. Chasman**[8,25], **Eco J.C. de Geus**[22,23,24], **Timothy M. Frayling**[16], **Joel N. Hirschhorn**[5,6,7,8], **Jouke Jan Hottenga**[22,23,24], **Erik Ingelsson**[11,26], **Ruth J.F. Loos**[27,28,29,30], **Patrik K. E. Magnusson**[31], **Nicholas G. Martin**[32], **Grant W. Montgomery**[32], **Kari E. North**[13,33], **Nancy L. Pedersen**[31], **Timothy D. Spector**[34], **Elizabeth K. Speliotes**[14], **Michael E. Goddard**[35,36], **Jian Yang**[1,9], and **Peter M. Visscher**[1,9]

[1]Queensland Brain Institute, The University of Queensland, Brisbane 4072, Australia [2]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands [3]Institute for Maternal and Child Health – IRCCS "Burlo Garofolo", Trieste, Italy [4]Department of Medical Statistics, University of Trieste, Italy [5]Estonian Genome Center, University of Tartu, Tartu 51010, Estonia [6]Divisions of Endocrinology and Genetics and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA 02115, USA [7]Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge 02142, MA, USA [8]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA [9]The University of Queensland Diamantina Institute, The University of Queensland, Translational Research Institute, Brisbane, QLD 4102, Australia [10]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA [11]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala 75185, Sweden [12]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA [13]Department of Internal Medicine, Division of Gastroenterology, and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA [14]Center for Statistical Genetics, Department of

Supplementary Information is linked to the online version of the paper at www.nature.com/ng and an online interactive version of Figure 1 can be found at http://www.kn3in.com/eu_traits/

Author contributions

Conceived and designed the study: Matthew R. Robinson, Michael E. Goddard, Jian Yang and Peter M. Visscher. *Data analysis:* Matthew R. Robinson with additional contributions from Gibran Hemani, Carolina Medina-Gomez, Massimo Mezzavilla, Konstantin Shakhbazov, Tonu Esko, Sonja I. Berndt, Stefan Gustafsson, Anne E. Justice, Bratati Kahali, Adam E. Locke, Tune H. Pers, Sailaja Vedantam, Andrew R. Wood, and Wouter van Rheenen. *Study oversight, sample collection, and management:* Jan H. Veldink, Ole A. Andreassen, Paolo Gasparini, Andres Metspalu, Fernando Rivadeneira, Thomas M. Werge, Goncalo R. Abecasis, Dorret I. Boomsma, Daniel I. Chasman, Eco J.C. de Geus, Timothy M. Frayling, Joel N. Hirschhorn, Jouke Jan. Hottenga, Erik Ingelsson, Ruth J.F. Loos, Patrik K.E. Magnusson, Nicholas G. Martin, Grant W. Montgomery, Kari E. North, Nancy L. Pedersen, Timothy D. Spector, Elizabeth K. Speliotes. *Wrote manuscript:* Matthew R. Robinson and Peter M. Visscher, with contributions from all authors on the final version.

The authors declare no conflicts of interest pertaining to this study.

Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA [15]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby 2800, Denmark [16]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter EX1 2LU, UK [17]Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands [18]NORMENT, KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway [19]Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Dervices Copenhagen, DK-4000 Roskilde, Denmark [20]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark [21]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark [22]Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, 1081 HV Amsterdam, The Netherlands [23]EMGO+ Institute for Health and Care Research, VU University Medical Center, 1081 BT Amsterdam, The Netherlands [24]Department of Biological Psychology, VU University Amsterdam, van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands [25]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215, USA [26]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK [27]MRC Epidemiology Unit, University of Cambridge, Institute of Metabolic Science, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ, UK [28]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA [29]The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA [30]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA [31]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-171 77 Stockholm, Sweden [32]QIMR Berghofer Medical Research Institute, 300 Herston Road, Q 4029, Australia [33]Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA [34]Department of Twin Research and Genetic Epidemiology, King's College London, St. Thomas' Hospital, London SE1 7EH, UK [35]Biosciences Research Division, Department of Primary Industries, Victoria 3083, Australia [36]Department of Food and Agricultural Systems, University of Melbourne, Victoria 3010, Australia

Abstract

Across-nation differences in the mean of complex traits such as obesity and stature are common[1–8], but the reasons for these differences are not known. Here, we find evidence that many independent loci of small effect combine to create population genetic differences in height and body mass index (BMI) in a sample of 9,416 individuals across 14 European countries. Using discovery data on over 250,000 individuals and unbiased estimates of effect sizes from 17,500 sib pairs, we estimate that 24% (95% CI: 9%, 41%) and 8% (95% CI: 4%, 16%) of the captured additive genetic variance for height and BMI across Europe are attributed to among-population genetic differences. Population genetic divergence differed significantly from that expected under a null model (P <3.94e^{-08} for height and P<5.95e^{-04} for BMI), and we find an among-population genetic correlation for tall and slender nations (r = −0.80 (95% CI: −0.95, −0.60), contrasting no genetic correlation between height and BMI within populations (r = −0.016, 95% CI: −0.041, 0.001), consistent with selection on height genes that also act to reduce BMI. Observations of

mean height across nations correlated with the predicted genetic means for height (r = 0.51, P<0.001), so that a proportion of observed differences in height within Europe reflect genetic factors. In contrast, observed mean BMI did not correlate with the genetic estimates (P<0.58), implying that genetic differentiation in BMI is masked by environmental differences across Europe.

Many of the phenotypes that vary within human populations are complex, in that they are determined by alleles at multiple loci as well as by many non-genetic factors[9–15]. Therefore, it is reasonable to assume that regional differences in such traits also have a complex basis[16–18]. Understanding these regional differences requires knowledge of the relative roles of environmental versus genetic effects, which can be gained through estimating the amount of population genetic variance in phenotype, and by determining the amount of observed differences that can be explained by population genetic effects[19]. However to date, these estimates have yet to be made outside of laboratory study populations[20], and experimental designs to estimate among-population differences due to genetic factors in human populations have been lacking due to the confounding of genetic and environmental effects.

At least 135 million European citizens are classed as obese[21], which is expected to have major direct and indirect health and economic costs[16, 22, 23]. Regional differences across Europe in height and susceptibility to weight gain, as defined by body-mass-index (BMI), are well documented[5, 18, 22–26], but the reasons for these differences are not well understood. Height and BMI are complex traits that have been extensively studied and there is strong evidence that both are influenced by a large number of genetic polymorphisms, with a significant proportion of their genetic variance captured by common SNPs[11, 26–30]. For height, there is strong empirical evidence for selection on height-associated single nucleotide polymorphism (SNP) loci within Europe[26, 31], and between European populations as compared to the rest of the world[32]. However, it is unlikely that the true extent of the population-genetic effects has been captured, or well represented by a limited number of ascertained loci examined to date. For BMI, we do not know whether genetic differentiation exists, and for both traits estimates of the contribution of common loci to population genetic variance and the observed regional differences remain unknown (although see[26]).

In this study, we estimate the cumulative population genetic differentiation at multiple unlinked loci associated with height and BMI across 9,416 European individuals from 14 countries, using population genetic analyses (Online Methods and Supplementary Figures S1 and S2). We performed GWAS meta-analyses on data from recent studies[33, 34], to select independent loci (r^2<0.1 and >1Mb distance using the PLINK clumping procedure[35]) that were associated with both traits in a large sample (~250,000 individuals for height and ~350,000 for BMI) of European ancestry. We then re-estimated the effects of each SNP in a within-family design, which is unbiased of population stratification, and used these effect size to create a genetic predictor for height and BMI (also termed ‘profile’ or ‘polygenic score’)[36]. The proportion of variance in the profile scores attributable to population differences was estimated in a Bayesian mixed effects model, alongside the co-differentiation of height and BMI, and the mean profile score for each nation (Online Methods and Supplementary Figure S1). Using theory and simulation study, we show that

whilst the use of within-family effect sizes provides unbiased estimated effect sizes, population genetic analyses conducted using loci that were ascertained from a standard GWAS can be biased if population stratification was not fully accounted for (Online Methods and Supplementary Figures S3, S4 and S5). In large-scale meta analyses, there is no certainty that population stratification has been completely controlled for and thus we repeated our analysis using (i) a non-ascertained set of unlinked (LD r^2<0.1 and >1Mb distance apart), common (minor allele frequency > 1%), HapMap3 loci (~40,000 SNP loci), and (ii) a set of unlinked (LD r^2<0.1 and >1Mb distance apart), common (minor allele frequency > 1%), HapMap3 loci selected based on their within-family association with each phenotype (~40,000 SNP loci for both traits). This provides genome-wide estimates of population genetic differentiation at common, unlinked loci, which are unbiased of population stratification or ascertainment biases, representing a lower-limit of the population-level effects.

Genetic differences among populations may occur by random chance processes, or through natural selection in our evolutionary past[19, 37–45]. Therefore, at all stages our results were compared to a null model representing the random chance process of genetic drift, which tests whether selection has acted on common variants to alter the frequency of height- and BMI-associated loci across populations. We estimate the population genetic co-differentiation of height and BMI across populations, to examine whether differentiation of both phenotypes is independent, which asks whether selection has acted on both traits independently. We then estimate whether population genetic effects for both phenotypes reflect the pattern of observed differentiation. Finally, we determine whether population-genetic effects are strongest at loci that are expected to explain the most variance in phenotype.

The maximum proportion of variance in a polygenic predictor (profile score) attributable to population differences was 24% (95% CI: 9%, 41%) and 8% (95% CI: 4%, 16%) for height and BMI, using 2,660 SNPs for height and 11,919 SNPs for BMI. For height, the largest proportion of population-level variance was captured by SNPs of the second lowest p-value threshold from the large-scale meta-analysis (Supplementary Figure S6). For BMI, the continual addition of SNPs increased the proportion of population-level variance (Supplementary Figure S6). These results reflect the fact that among-population variation is greater in a predictor explaining a greater proportion of the phenotypic variance within a population for both phenotypes (Supplementary Figure S7) and thus the estimates presented here may increase as prediction accuracy increases in future. Our results were confirmed in both the non-ascertained set of independent genome-wide loci, (for height: 8.6%, 95% CI 3%, 15.7%; and for BMI: 2.8%, 95% CI 1.1%, 5.3%), and the independent set of loci selected on their association with both traits in the within-family analysis (for height: 11.9%, 95% CI 4.5%, 21.8%; and for BMI 8% 95% CI: 3.4%, 14.7%). The lower among-population variance at the non-ascertained set of loci, reflects the fact that the prediction accuracy was reduced, likely through the addition of a very large number of loci with no detectable association with either phenotype. Subsequent results are presented using the predictor that captures the greatest amount of population- and individual-level variance (2,660 SNPs for height and 11,919 SNPs for BMI); however, the pattern of population means remained the same at both sets of genome-wide loci (Supplementary Figure S8).

Model estimates of the predicted population genetic means for height and BMI are given in Figure 1 alongside the observed values which we estimate from an independent set of recently published data25,46, whilst accounting for time trends. Population genetic differences in allele frequency are expected to create genetic differences in height such that people of the Netherlands are on average 1cm taller than those in Italy, and create genetic differences in BMI such that on average people from Italy and Denmark differ by 0.2 BMI units (Figure 1 and Figure 2).

We next determined whether we could reject the null hypothesis that the population genetic differentiation observed for height and BMI reflected a pattern expected under neutrality. To do this, we compared our estimates to a null quantitative genetic model of multivariate population differentiation[32, 47]. We found strong evidence that the divergence of each trait was greater than would be expected under a neutral model (Figure 2). The overall level of neutral genetic differentiation was comparatively low and estimates of the population-level variance expected under drift were small for both height 1.2% (0.01%, 1.78% 95%CI) and BMI 1.9% (0.48%, 2.97 95%CI), and not significantly different to the average F_{ST} of the SNP sets between the populations of 1% for height and 1.2% for BMI. Our results were confirmed in both the non-ascertained set of independent genome-wide loci, (For height: $p = 3.29\times10^{-06}$; and for BMI: p=0.018), and an independent set of loci selected on their association with both traits in the within-family analysis (For height: $p = 2.67\times10^{-06}$; and for BMI p = 8.35×10-05). We therefore reject the null model and our results suggest that population genetic differentiation across these 14 European countries for height and BMI have been driven by selection on standing genetic variation across geographical regions in our evolutionary past. The significant departure from a neutral model occurs because on average, the common loci comprising the genetic predictor are differentiated in a direction that is consistent with the direction of their effects on each trait, which in turn creates differences among countries in a genetic predictor that are greater than expected by chance.

Our rejection of the null hypothesis, that population differences in polygenic score for height and BMI are due to drift is not caused by potential SNP ascertainment biases or population stratification. Our interpretation that the differences in polygenic score between populations is due to selection for height and BMI does not depend on the LD between causal variants because we have used only independent loci; nor does it depend the SNPs used because we generate our null model estimates from the same set of SNPs that we use to create the genetic predictor (see Online Methods). Loss of LD, genotype-environment interactions, and genetic heterogeneity, can change the relationship between SNPs and the underlying causal variants across countries. However, this would reduce the likelihood of detecting population genetic differentiation in our approach and could not cause us to incorrectly reject our null hypothesis. In this study, we minimize these effects because profile scores were created from SNPs that had their effect sizes estimated within-populations of European ancestry.

Our framework provides an estimate of the co-divergence in population mean genetic profile scores between height and BMI (Figure 2), and we found a negative correlation among population genetic means of −0.80 (95% CI: −0.95, −0.60), which was consistent across predictors made genome-wide sets of loci (for genome-wide loci selected based on within-family association with each trait: −0.89, 95% CI −0.97, −0.77; and for non-ascertained

genome-wide loci: −0.77, 95% CI −0.94, −0.55). This result implies selection acting on common loci in a way that increases height whilst reducing BMI and visa versa. Thus, we find that across the majority of the European countries a genetic predisposition to being a tall nation was associated with a genetic predisposition to being a slender nation (low BMI). Our model also enabled us to estimate the correlation among genetic profile scores within-populations (Online Methods), which yielded an estimate of r = −0.016 (95% CI: −0.041, 0.001). The fact that height and BMI are nearly uncorrelated at the individual-level means that selection acting on one trait would not be expected to elicit a response in the other, suggesting that selection has acted on both height and BMI-associated loci. We cannot rule out genetic differentiation in one trait being mediated to some extent by selection upon the other as some genes affect both phenotypes[48]. However, our results do strongly suggest that the population genetic co-divergence shown here is inconsistent with random genetic drift, because under drift the expectation is that the among-population genetic correlation should equal the within-population correlation[47, 49].

We then tested if the phenotypic differences across the 14 European countries observed today reflect genetic differentiation at common height- and BMI-associated loci, or whether current environmental differences among countries (diet, economy, climate etc.) mask the population genetic differentiation that we detect. Our results show a strong association (r=0.51; 95% CI 0.39, 0.61; P<0.001) between the population genetic values and the observed phenotypic pattern for height (Figure 3). This suggests that the phenotypic height differences that we observe across countries are partially due to differences in allele frequency at common height-associated loci. For BMI, the pattern of population genetic differentiation did not reflect the observed pattern (r = −0.10; 95% CI −0.19, 0.01; P = 0.584; Figure 3). This suggests that, although selection has created population genetic differentiation for BMI, environmental differences among countries mask population genetic differentiation. We also estimated the correlation between the phenotypic differentiation expected under drift and the observed values for both phenotypes and found no evidence of a correlation (Figure 3), implying that drift does not drive the observed national patterns of either trait.

We identified loci that contributed most to the genome-wide pattern of population genetic differentiation that we estimated for height and BMI (see Online Methods). We found greater differentiation in allele frequency across populations at SNPs with intermediate allele frequency, and an association between the expected additive genetic variance contributed by a SNP (estimated from the within-family effect size estimate and the SNP allele frequency) and its contribution to population genetic differentiation (Supplementary Figure S9 and Supplementary Figure S10), a pattern expected from our theory (Supplementary Figures S11 and S12). This suggests that SNPs making the largest contribution to the phenotypic variance are enriched for association with the genome-wide pattern of population genetic differentiation, and this is evidenced by the fact that the proportion of population-level variance is greater in a predictor explaining a greater proportion of the phenotypic variance for both phenotypes. We found no evidence for significant population differentiation at any SNP at a genome-wide threshold of $p<7\times10^{-7}$ within our European prediction sample, implying that many loci of small effect across the genome combine to create the genome-wide population genetic differentiation that we

detect. Annotation of the top 500 contributing SNPs of each trait to genes followed by enrichment analyses, suggest that population genetic variation across the 14 countries for height and BMI is likely underlain by the combined effects of multiple pathways across the genome, with overlap in the genes involved (20% overlap of annotated genes, Supplementary Figure S13).

Finally, we examined population genetic differentiation in height and BMI at both a local and a world-wide scale. We found no evidence for population genetic differentiation across six Northern Italian villages[50] (Supplementary Figure S14), suggesting that genetic differentiation among subgroups of such a genetically homogeneous population are likely to be small. We then examined population differentiation for height and BMI in the Human Genetic Diversity Panel (HGDP), as used in a previous study[32] (Supplementary Figure S15). We found evidence to reject the null hypothesis that population genetic differentiation observed for height and BMI reflected a pattern expected under neutrality at genome-wide independent loci (Supplementary Figure S15). This extends previous work[32], which reported no significant differentiation for BMI using a limited number of loci and we also find no evidence for population genetic co-differentiation of height and BMI, implying a pattern of selection that is specific to Europe.

The conclusions of our study are fourfold: (i) many common loci combine in a consistent manner to create population genetic differences for height and BMI; (ii) we reject our null hypothesis that the population genetic differentiation for height and BMI reflects a pattern expected under neutrality and thus conclude that selection drives the population genetic differentiation that we observed; (iii) population genetic differentiation for height and BMI are correlated within Europe, which is also to a greater level than expected under neutrality, implying that height loci under selection are enriched for effects that reduce BMI; (iv) the selection driven population genetic differences for height are associated with the phenotypic patterns we see across Europe. The selection driven population genetic differences for BMI are not associated with the phenotypic patterns and thus we conclude that environmental factors are masking the population genetic differences. While, genotype-environment effects and rare variants will play a role in shaping population genetic differentiation, the focus of this approach is on estimating the amount of population-level variance that is ‘tagged’ by a specific set of common SNP markers. As additional genetic variation is captured for both traits, it is likely that power will increase to fully capture the among population genetic effects. The theoretical and analysis framework builds upon previous approaches[26, 32], is entirely general, and can be applied to estimate the role of commonly varying loci in shaping population differences in any set of phenotypes.

ONLINE METHODS

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:

$$\sigma^2_{p,k}=(\frac{1}{n_r})\sum_{r=1...n_r}(p_{k,r}-\overline{p}_k)^2 \quad [1.1]$$

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:

$$\theta_k=\frac{\sigma^2_{p,k}}{\overline{p}_k(1-\overline{p}_k)} \quad [1.2]$$

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:

$$\sigma^2_{g,k}=2\overline{p}_k(1-\overline{p}_k)a^2_k \quad [1.3]$$

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:

$$\sigma^2_{r,k}=2\overline{p}_k(1-\overline{p}_k)a^2_k\theta_k \quad [1.4]$$

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:

$$Q=\frac{\sigma^2_{r,k}}{\sigma^2_{g,k}}=\theta \quad [1.5]$$

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 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:

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

and the genetic variance across populations is then:

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

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

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

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:

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

Additive variance at the population level is then:

$$\sigma^2_g=(\frac{1}{n_r})\sum\nolimits_{X\in r}(\overline{g}_r-\overline{g})(\overline{g}_r-\overline{g})^T, \quad [1.11]$$

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

$$Q=\frac{\sigma^2_{\mathrm{gr}}}{\sigma^2_g} \quad [1.12]$$

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:

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

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 $\hat{g}_t = \Sigma_k x_{k,i} \hat{\beta}_{k,m}$, 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:

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

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

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

$$\hat{\boldsymbol{g}}=\boldsymbol{\mu}+\boldsymbol{v}+\boldsymbol{e}, \quad [2.2]$$

where $\hat{g}$ is a matrix of genetic predictors of individual i for traits m, μ 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 $\bar{p}_k = 0.5(p_{1k} + p_{2k})$, where p_{1k} and p_{2k} are the frequencies of SNP k in populations 1 and 2 respectively. We define the difference in allele frequency among populations as $\delta_k = p_{1k} - p_{2k}$ and $F_{\mathrm{ST}_K}=\frac{0.5(p_{1_k}+p_{2_k})^2}{\overline{p}k(1-\overline{p}k)}$ We denote the population means as $\bar{y}_1$ and $\bar{y}_2$. We define the difference in the population means as $\bar{y}_1 - \bar{y}_2 = 2\beta(p_{1k} - p_{2k}) + \Delta$, 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:

$$\hat{\beta}_k=\frac{\sigma(y,x_k)}{\sigma^2_{x_k}},$$
$$\sigma^2_{x_k}=E(x_k^2)-E(x_k)^2=2\overline{p}_k(1-\overline{p}_k)(1-0.5F_{\mathrm{ST}_k}),$$

dropping subscript k for convenience. Since,

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

then,

$$\begin{aligned}\sigma(x,y)&=E(xy)-E(x)E(y),\\&=\beta(p_1(1-p_1)+p_2(1-p_2))+\tfrac{1}{2}(\overline{y}_1-\overline{y}_2)(p_1-p_2),\\&=\beta(p_1(1-p_1)+p_2(1-p_2))+\tfrac{1}{2}(p_1-p_2)(2b(p_1-p_2)+\Delta),\\&=b(2\overline{p}(1-\overline{p}))\left(1+\tfrac{1}{2}F_{\mathrm{ST}}\right)+\tfrac{1}{2}\Delta(p_1-p_2),\end{aligned}$$

and hence,

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

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

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:

$$E(T\backslash\delta)=2\overline{p}(1-\overline{p})N\beta_\delta^2 \quad [2.4]$$

$$\sigma^2(T\backslash\delta)=2+4\overline{p}(1-\overline{p})N\beta_\delta^2 \quad [2.5]$$

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

$$\begin{aligned}E_\delta(\beta_\delta^2) &=E(\beta+\tfrac{1}{4}\Delta(\overline{p}(1-\overline{p}))(p_1-p_2))^2,\\&=\beta^2+(\overline{y}_1-\overline{y}_2)^2F_{\mathrm{ST}}/8\overline{p}(1-\overline{p}) \quad [2.6]\end{aligned}$$

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

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

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

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

Scaled by N, the mean and variance of the NCP are $q^2 + ¼N(\bar{y}_1 - \bar{y}_2)^2\, F_{ST}$ and $2q^2 + ½N(\bar{y}_1 - \bar{y}_2)^2\, F_{ST}$, 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} ~ 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 $(\overline{y}_1 - \overline{y}_2)r\sqrt{\frac{\sigma^2_U}{\sigma^2_{x_k}}}$, 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:

$$\hat{g}_i = \sum\nolimits_k x_{k,i}\hat{\beta}_k + \sum\nolimits_k \frac{\frac{1}{2}\Delta(p_{1_k} - p_{2_k})}{2p_k(1-p_k)(1+\frac{1}{2}F_{\mathrm{ST}_k})} \quad [2.9]$$

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 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[55]. For sibling pairs, the model is:

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

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 $\Sigma x_{k,f}$ 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:

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

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:

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

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.

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

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:

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

where μ is the mean additive genotype in the ancestral population which is determined by the allele frequencies in the ancestral population, **I** is a unit vector relating observations to populations, $\mathbf{G}^{\mathrm{A}}$ is the ancestral variance-covariance matrix of the traits in question, Θ^r is the matrix of population-level coancestry, and ⊗ 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 Θ^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.

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 Σ.

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 $D=\sqrt{(\boldsymbol{A}^r-\boldsymbol{A}^D)^T\sum^{-1}(\boldsymbol{A}^r-\boldsymbol{A}^D)}$. There is a single mean estimate of Σ, 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 height and BMI associated loci that contribute to the population genetic effects.

In Eq 4.2, $\overline{g}_{1_k}-\overline{g}_{2_k}\sim N(0,\sigma^2_\delta)$ and by standardizing the predictor of each SNP k to a variance of 1, $\bar{g}_{1k}-\bar{g}_{2k}\sim N(0,1^2)$. 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, $\bar{g}_k$ and a vector of standardized population effects, v_r, gives a standard normal density:

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

where δ_k^2 follows 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 $\mathbf{H^o}$ is calculated as

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

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 $\mathbf{H^p}$ for the profile scores of the quantitative traits can also be defined as

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

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

The association between $\mathbf{H^o}$ to $\mathbf{H^P}$ can be determined using a Mantel test statistic[54]:

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

where n_r denotes the number of populations.

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

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

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

Parameter uncertainty is accounted for because H is calculated for each recorded MCMC chain t, with the fraction of cases where $\mathrm{M}(\mathbf{H}^o,\mathbf{H}_t^P)>\mathrm{M}(\mathbf{H}^o,\mathbf{H}_t^{P,R})$ 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 $\mathbf{H}^{\boldsymbol{o}}$ to $\mathbf{H}^{\boldsymbol{P}}$ to be compared on the same SD scale.

6. Simulation study of estimation and prediction bias

Approximations of the parameters—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:

$$\sigma^2_{r,k}{=}2\overline{p}_k(1-\overline{p}_k)a_k^2\theta_k \quad [6.1]$$

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:

$$a_k{\sim}N(0,\sqrt{\frac{h^2}{2p_k(1-p_k)n_k}}) \quad [6.2]$$

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, 56].

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

$$\theta_j{=}p_k(1-p_k)\mathrm{Beta}(\alpha,\beta) \quad [6.3]$$

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[57, 58].

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

$$p_{k,r}=p_k+\theta_k \otimes S_r \quad [6.4]$$

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:

$$x_{k,i,r}\sim \mathrm{Bin}(p_{k,r,}2) \quad [6.5]$$

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:

$$\hat{g}_{i,r}=\sum_k x_{k,i,r}\alpha_k \quad [6.6]$$

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:

$$\hat{g}_i=\mu+\upsilon_r+e_i, \quad [6.7]$$

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:

$$g_{i,r}=\sum_k x_{k,i,r}(\alpha_{k,r,m}+\in_{k,r,m}), \in_{k,r,m}=N(0,z) \quad [6.8]$$

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

Eq. 6.1 through Eq. 6.8 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_i; and the cumulative error variance within the profile score can be modified by altering z in Eq. 6.8. 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:

$$\sigma^2_{A,k}=2\overline{p}_k(1-\overline{p}_k)a^2_k \quad [6.9]$$

We performed simulations using Eq. 6.1 through Eq. 6.9. 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.

Simulation study using real genotype data—We used the common, independent, HapMap3 SNPs from the 17,500 sibling pairs used in the main empirical analyses (Supplementary Table 1) as the basis of a series of simulations. Causal variants were allocated to 5000 of the independent loci at random across the genome, with their effects sampled from a normal distribution, with mean of 0 and variance 1. The heritability of the trait was simulated to be 90% and a phenotype was created as $y = \Sigma x_k b_k + e$, with e = N(0,1-h^2). 50 simulation replicates were conducted.

For each simulation replicate, we randomly selected 16,000 sibling pairs to create an estimation set, leaving 1,500 pairs as a prediction set. We then tested the effects of each SNP on the phenotype in the estimation set in three ways. First, we used a within-family sibling pair analysis implemented as the QFAM procedure in PLINK described in Eq. 3.1. Second, we selected one member of a sibling pair at random to create an unrelated set of individuals and then estimated SNP effects in a ordinary least squares regression (the standard GWAS approach) without any control for population stratification. Finally, we repeated the GWAS estimation controlling for the first 20 principal components estimated from the sibling pair sample.

We then used these three sets of estimates to create three different profile scores in the prediction set. We followed a recent approach to partition variance in a predictor in sibling pairs into genetic, environment, and common genetic and common environment terms[33]. Variance attributable to common genetic or common environment terms indicates population

stratification bias in the effect size estimates, enabling us to demonstrate that our within-family estimates are unbiased of population stratification.

Additionally, we used the three sets of effect size estimates to create three different profile scores in the European prediction data of 9416 individuals used in the main empirical analyses. Our European prediction data was projected onto the first principal component estimated in the within-family sample, and then two groups of individuals were created based on the upper and lower quartiles of the distribution of the projected principal component. The first principal component has been shown to reflect population stratification within genotype samples and thus, we stratify the independent prediction sample by the predominant axis of potential bias in the discovery sample. If there is no population stratification bias in the estimates of the SNP effects then there should be no significant differentiation between the upper and lower quartile groups in the mean of the genetic predictor. We therefore compare the estimates gained from the three predictors to those made when a predictor is created using the true simulated effects, and to those made when a predictor is created using randomly allocated effect sizes under our null model. To test for ascertainment biases, we selected the top 100 and top 500 SNPs identified by a GWAS that does not control for population stratification, created two predictors at these SNPs, one from the GWAS effect size estimates and one from the within-family estimates, and then compared the mean differences between the upper and lower quartile groups. This identifies SNPs from a GWAS and then tests for population differentiation as in previous approaches.

Finally, we repeated the 50 simulations described above, but we created a genotype-environment correlation. We used the same effect sizes and phenotype as in the simulation described above, but we added a z-score of the standardized individual-level eigenvalue of the first principal component estimated in the within-family data. Thus our phenotype was: $y = \Sigma x_k b_k + 0.2PC_1 + e$, where PC1 is the z-score standardized value for each individual at the first principal component. This creates a correlation between the phenotype and PC1 of ~0.2 across simulations, representing a phenotypic difference that aligns to the major axis of genetic differentiation. Note that in this scenario there is no selection because causal variants are allocated at random across the genome and so on average across simulations their frequency is not expected to differ in a manner that will create a consistent directional difference in profile score along any axis of population stratification. We repeated the estimation and prediction into our independent European prediction sample to test for a directional deviation in mean profile score between the upper and lower quartile groups. Under this scenario ascertainment bias are expected to be large because SNPs identified by a GWAS that does not control for population stratification should be those with the strongest genotype-environment correlation, creating a predictor with mean differences across the leading principal component. However, genome-wide the expectation is that the direction of SNP differentiation should not align with the direction of the effect size and the direction of the phenotypic differentiation, provided that predictors are created from SNP effect sizes that are unbiased of population stratification.

7. Differentiation of height and body mass index across Europe

We apply this framework to examine population differentiation in height and body mass index (BMI) across Europe. Our analysis is a sequence of steps (Supplementary Figure S1): (1) quantify the association of genome-wide SNPs with height and BMI using large-scale meta-analyses; (2) re-estimate the SNP effects in a within-family design that is unbiased of population stratification; (3) use the within-family estimates to predict the genetic value (create a profile score) of individuals in an independent data set and obtain a genome-wide predictor with greatest predictive power for both phenotypes; (4) predict height and BMI across a number of populations from genomic data; (5) partition variance in genetic value into population- and individual-level effects and estimate the population-level covariance across traits to test for correlated population differentiation; (6) test whether the pattern of population-level (co)variance reflects a signal of selection or that expected by drift using a null model; (7) examine the amount of observed phenotypic differentiation that can be explained by genetic differentiation across populations; (8) confirm the results by creating a predictor by combining a non-ascertained set of genome-wide, common, independent, HapMap3 SNPs, and their within-family effect sizes that are unbiased of population stratification; and (9) identify the leading SNPs driving the genome-wide pattern of population genetic differentiation and test for a relationship between phenotypic variance explained by a locus and its association with population genetic differentiation.

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: (f) 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, 59] ; (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].

Imputation—All of the cohorts used in this study were independently imputed to the 1000 genomes reference panel, using identical QC procedures on the initial datasets of per-SNP missing data rate of <0.02, minor allele frequency >0.01, per-person missing data rate <0.05, and Hardy-Weinberg disequilibrium p-value 0.001. Imputation for the majority of cohorts was performed in two stages. First, the target data was haplotyped using HAPI-UR[60]. Second, Impute2[61] was used to impute the haplotypes to the 1000 genomes reference panel[62] (release 1, version 3). We then selected SNPs which were present across all datasets at an imputation information score of >0.8. A full imputation procedure is described at https://github.com/CNSGenomics/impute-pipe. The imputation for the Netherlans cohort was identical, except SHAPEIT2[63] was used for haplotyping. We performed these same QC steps again after combining data from different cohorts, including comparisons of allele frequencies across populations.

Selection of SNPs for genomic profiling across Europe—We performed GWAS meta-analyses on data from recent studies[33, 34], to select independent loci (r^2<0.1 and >lMb distance using the PLINK clumping procedure[35]) that were associated with both traits in a large sample (~250,000 for height and ~350,000 for BMI) individuals of European ancestry. We then re-estimated SNP effects at these loci in a within-family sibling pair dataset (Supplementary Table 1) using the QFAM procedure in PLINK described in Eq 3.1. Using an independent set of data (Supplementary Table 1), we then identified the set of loci that when combined with the within-family effect sizes into a predictor, captured the largest amount of phenotypic variance for each phenotype.

We then confirmed our results using a non-ascertained genome-wide set of unlinked (LD r^2<0.1 and >lMb distance apart), common (minor allele frequency > 1%), HapMap3 loci, with height and BMI (~40,000 SNP loci) that passed QC in both the within-family and the prediction sample. We estimated the effects of these SNPs again using our within-family sibling pair dataset using the QFAM procedure in PLINK described in Eq 3.1.

Genomic profile scoring in a collection of European genomic data—We used these within-family effect sizes to create genomic profile scores for individuals across a collection of European genomics data. All data was imputed as described above, and details of the cohort are provided in Supplementary Table 1. From the POPRES study we selected individuals from France, Portugal, Spain, Italy and Switzerland whose grandparents were born in the same country as the sample individuals. From the Estonian and Finnish cohorts we selected 1000 individuals at random that were used for all analyses. For the Netherlands cohort, we selected 1000 individuals from the MinE ALS study (www.projectmine.com). These individuals were healthy controls, born in the Netherlands, whose grandparents and parents were also born in the Netherlands. From the Psychiatric genomic consortium and the Welcome Trust Case Control Consortium 2, we used control individuals from Bulgaria, Ireland, Norway, Denmark, Sweden and the UK (Supplementary Table 1). Eq. 2.2 was estimated using the R package MCMCglmm[64], with uninformative inverse Wishart priors, a

burn in period of 7,000 iterations, a sampling interval of 10 iterations, and a total number of iterations of 17,000, providing 1000 posterior estimates.

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:

$$w_{i,m}=\mathrm{SD}(m)\mathrm{cov}(z_y, z_{\hat{g}})z_{\hat{g}} \quad [7.1]$$

where 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 $g_{\wedge i,m.}$

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.

We used an independent sample of population data to determine the amount of phenotypic variance explained by our profile score, as measured by the $cov(z_y,Z_{\hat{g}})$ in Eq 7.1. Individuals within the Health and Retirement Study (Supplementary Table 1) were unrelated, and the phenotype was adjusted by the first 20 principal components of the SNPs used in the predictor, to account for any population stratification before estimating the within-population covariance. The SD of height and BMI measured in the sample of 17,500 quasi-independent sib-pairs were estimated accounting for sex differences. Eq. 7.1 was then used to transform the profile scores onto the observed scale, and we present these graphically within the figures.

European phenotypic data—BMI and height measures for males for each of the 14 European countries were taken from recently published estimates[25, 46]. For BMI, measures were available from 1980 to the present day, and for height measures were available from 1860 to the present day. Both phenotypes were adjusted for time trends before estimating the population means within a mixed effects model. The model estimates of these means are then compared to the predicted genetic means as described above as described in Eq 5.1 to 5.4.

Testing the contribution of each SNP to the pattern of population differentiation—As described above in Eq. 4.4, we estimated a χ^2 value for each SNP and tested for the association between a SNPs contribution to differentiation and its minor allele

frequency, and expected contribution to phenotypic variation estimated as $2pq\hat{\beta}^2$ using the within-family effect sizes. We then performed GWAS meta-analyses on data from recent studies[33'34], excluding cohorts present in our within-family and prediction samples. We excluded a number of cohorts that were used in the GIANT meta-analyses. For height, we excluded the Netherlands Twin Gene study, Twin UK, the Queensland Institute of Medical Research (QIMR) sample, the Framingham Study sample, and the Netherlands Twin Register. For BMI, cohort-level summaries were not available for all samples and we could only exclude the QIMR sample.

Positional and functional annotation of SNPs—Our aim in this section was to describe the positional genic annotation and gene ontology categories for height-and BMI-associated SNPs that contribute most to the genome-wide pattern of population genetic variance. We select the top 500 differentiated height and BMI SNPs, and did three things:

- **i.** based on the genomic position of these SNPs we assigned them to genes and we simply estimated the number of overlapping genes involved in the genetic differentiation of height and BMI.
- **ii.** based on the genomic position of these SNPs we assigned positional genic annotation categories of: (1) 3' untranslated region (3'UTR); (2) 5' untranslated region (5' UTR); (3) intronic variants; (4) non-coding transcription variants; (5) 1 to 1,000 base pairs downstream; (6) 1 to 1,000 base pairs upstream; (7) missense transcription variants; (8)synonymous transcription variants; (9) non-coding exonic variants.
- **iii.** based on the genomic position of these SNPs we assigned them to Ensembl gene identities and then to gene ontology (GO) terms.

We then conducted statistical testing for parts (ii) and (iii). As a baseline, we used the top 10,000 SNPs for height and BMI. We then repeated steps (i) through (iii) using this larger set of SNPs. We first compared count data of the number of SNPs within each genic category from the highly differentiated SNPs, to count data from he top 10,000 SNPs. This provides a list of potentially enriched genic categories of highly differentiated SNPs. We used Fischer's exact tests (hypergeometric test), with Bonferroni *P*-value correction. Second, we then compared count data of the number of SNPs within each GO term from the highly differentiated SNPs, to count data from the top 10,000 SNPs. Again, we used Fischer's exact tests, with Bonferroni *P*-value correction. From this analysis, we selected a list of the top 20 potentially enriched functional categories for each trait. We assign *P*-values to our comparisons of count data, but use these only as a guide to select the top categories, rather than a definitive test of enrichment. All annotation was conducted using the R library biomaRt from Bioconductor (www.bioconductor.org).

Genomic profile scoring worldwide—We repeated our analyses for height and BMI using data from the Human Genome Diversity Panel as analyzed in[32]. We imputed the data following our protocol outlined above.

Code availability—Full computer code for the analysis and the values used to produce the figures is available from the lead author.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank three reviewers for very helpful and insightful comments that greatly improved the manuscript. We also thank Bill Hill and Otso Ovaskainen for useful discussions. The University of Queensland group is supported by the Australian National Health and Medical Research Council (NHMRC grants 1078037, 1048853 and 1050218). J.E.P. is supported by Australian Research Council grant DE130100691. J.Y. is supported by a Charles and Sylvia Viertel Senior Medical Research Fellowship and by NHMRC grant 1052684. We thank our colleagues at the Centre for Neurogenetics and Statistical Genomics for comments and suggestions. We are grateful to the twins and their families for their generous participation in the full-sibling family data, which includes many cohorts and received support from many funding bodies. The Swedish Ministry for Higher Education financially supports the STR. TWINGENE was supported by the Swedish Research Council (M-2005–1112), GenomEUtwin (EU/QLRT-2001-01254;QLG2-CT-2002-01254), NIH DK U01-066134, The Swedish Foundation for Strategic Research(SSF), and the Heart and Lung Foundation no. 20070481. Netherlands Twin Registry (NTR): Funding was obtained from the Netherlands Organization for Scientific Research (NWO:MagW/ZonMW grants 904-61-090, 985-10-002, 904-61-193, 480-04-004, 400-05-717, Addiction-31160008, Middelgroot-11-09-032, Spinozapremie 56-464-14192), Center for Medical Systems Biology (CSMB, NWO Genomics), NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI–NL, 184.021.007), the VU University's Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam (NCA), the European Science Foundation (ESF, EU/QLRT-2001-01254), the European Community's Seventh Framework Program (FP7/2007–2013), ENGAGE (HEALTH-F4-2007-201413); the European Science Council (ERC Advanced, 230374), Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the Avera Institute, Sioux Falls, South Dakota (USA), and the National Institutes of Health (NIH, R01D0042157-01A, Grand Opportunity grants 1RC2 MH089951-01 and 1RC2 MH089995-01). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health. The TwinsUK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project grant agreement (HEALTH-F4-2007-201413). The study also receives support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. TDS is an NIHR senior Investigator and is holder of an ERC Advanced Principal Investigator award. Genotyping for the TwinsUK study was performed by The Wellcome Trust Sanger Institute, support of the National Eye Institute via an NIH/CIDR genotyping project. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract N01-HC-25195). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI. Funding for SHARe Affymetrix genotyping was provided by NHLBI Contract N02-HL-64278. SHARe Illumina genotyping was provided under an agreement between Illumina and Boston University. The QIMR researchers acknowledge funding from the Australian National Health and Medical Research Council (grants 241944, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 496688 and 552485), and the National Institutes of Health (grants AA07535, AA10248, AA014041, AA13320, AA13321, AA13326 and DA12854). We are grateful to Michael Gill from Discipline of Psychiatry, School of Medicine, Trinity College Dublin and Kristin Nicodemus of the Institute of Genetics and Molecular Medicine, University of Edinburgh, UK for access to the ISC-Trinity College Dublin cohort, which was supported by the Welcome Trust and Health Research Board, Ireland. Access to the Bulgarian cohort data was kindly facilitated by George Kirov and Valentine Excott-Price. For the Danish cohort, the Danish Scientific Committees and the Danish Data Protection Agency approved the study and all the patients have give written informed consent prior to inclusion into the project. The National Institute of Aging (NIA) provided funding for the Health and Retirement Study (HRS) (U01 AG09740. The HRS is performed at the Institute for Social Research, University of Michigan. This manuscript was not prepared in collaboration with investigators of the HRS and does not necessarily reflect the opinions or views of the HRS, University of Michigan, or NIA. The HRS dataset and the POPRES dataset used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through accession numbers phs000428.v1.p1 and phs000145.v4.p2. The Netherlands genotype samples were part of Project MinE (www.projectmine.com), which was supported by the ALS Foundation Netherlands. Research leading to these results has received funding from the European Community's Health Seventh Framework Programme (FP7/2007-2013).

References

1. Moussavi S, et al. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet. 2007; 370:851–858. [PubMed: 17826170]

2. Wild S, Roglic G, Green A, Sicree R, King H. Global Prevalence of Diabetes: Estimates for the year 2000 and projections for 2030. Diabetes Care. 2004; 27:1047–1053. [PubMed: 15111519]
3. Dye C. Global Burden of Tuberculosis: Estimated Incidence, Prevalence, and Mortality by Country. JAMA. 1999; 282:677. [PubMed: 10517722]
4. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet. 2006; 367:1747–1757. [PubMed: 16731270]
5. Wang H, et al. Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012; 380:2071–2094. [PubMed: 23245603]
6. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol. Biomarkers Prev. 2010; 19:1893–1907. [PubMed: 20647400]
7. Kim AS, Johnston SC. Global variation in the relative burden of stroke and ischemic heart disease. Circulation. 2011; 124:314–323. [PubMed: 21730306]
8. Johnston SC, Mendis S, Mathers CD. Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling. Lancet Neurol. 2009; 8:345–354. [PubMed: 19233730]
9. Yang J, Visscher PM, Wray NR. Sporadic cases are the norm for complex disease. Eur. J. Hum. Genet. 2010; 18:1039–1043. [PubMed: 19826454]
10. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 2008; 4:el000008.
11. Yang J, et al. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 2010; 42:565–569. [PubMed: 20562875]
12. Morris AP, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. 2012; 44:981–990. [PubMed: 22885922]
13. Lee SH, et al. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. Hum. Mol. Genet. 2013; 22:832–841. [PubMed: 23193196]
14. Yang J, et al. Ubiquitous polygenicity of human complex traits: genome-wide analysis of 49 traits in Koreans. PLoS Genet. 2013; 9:el003355.
15. Robinson MR, Wray NR, Visscher PM. Explaining additional genetic variation in complex traits. Trends Genet. 2014; 30:124–132. [PubMed: 24629526]
16. Abegunde D0, Mathers CD, Adam T, Ortegon M, Strong K. The burden and costs of chronic diseases in low-income and middle-income countries. Lancet. 2007; 370:1929–1938. [PubMed: 18063029]
17. Kim AS, Johnston SC. Temporal and geographic trends in the global stroke epidemic. Stroke. 2013; 44:S123–S125. [PubMed: 23709707]
18. Ezzati M, Riboli E. Can noncommunicable diseases be prevented? Lessons from studies of populations and individuals. Science. 2012; 337:1482–1487. [PubMed: 22997325]
19. Hartl, DL.; Clark, AG. Principles of Population Genetics. Vol. 542. Sinauer Associates, Incorporated; 1997.
20. Leinonen T, McCairns RJS, O'Hara RB, Merila J. Q(ST)-F(ST) comparisons: evolutionary and ecological insights from genomic heterogeneity. Nat. Rev. Genet. 2013; 14:179–190. [PubMed: 23381120]
21. James TP, Rigby N, Leach R. The obesity epidemic, metabolic syndrome and future prevention strategies. Eur. J. Cardiovasc. Prev. Rehabil. 2004; 11:3–8. [PubMed: 15167200]
22. Popkin BM. Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. Am J Clin Nutr. 2006; 84:289–298. [PubMed: 16895874]
23. Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. Health and economic burden of the projected obesity trends in the USA and the UK. Lancet. 2011; 378:815–825. [PubMed: 21872750]

24. Ng M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014
25. Finucane MM, et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9-1 million participants. Lancet. 2011; 377:557–567. [PubMed: 21295846]
26. Turchin MC, et al. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat. Genet. 2012; 44:1015–1019. [PubMed: 22902787]
27. Speliotes EK, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat. Genet. 2010; 42:937–948. [PubMed: 20935630]
28. Lango Allen H, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467:832–838. [PubMed: 20881960]
29. Yang J, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 2012; 44:369–375. [PubMed: 22426310]
30. Yang J, et al. FTO genotype is associated with phenotypic variability of body mass index. Nature. 2012; 490:267–272. [PubMed: 22982992]
31. Amato R, Miele G, Monticelli A, Cocozza S. Signs of selective pressure on genetic variants affecting human height. PLoS One. 2011; 6:e27588. [PubMed: 22096598]
32. Berg JJ, Coop G. A Population Genetic Signal of Polygenic Adaptation. PLoS Genet. 2014; 10:el004412.
33. Wood AR, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 2014; 46:1173–1186. [PubMed: 25282103]
34. Locke AE, et al. Genetic studies of body mass index yeild new insights for obesity biology. Nature. 518:197–206. [PubMed: 25673413]
35. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 2007; 81:559–575. [PubMed: 17701901]
36. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013; 9:el003348.
37. Sabeti PC, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007; 449:913–918. [PubMed: 17943131]
38. Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. Recent and ongoing selection in the human genome. Nat. Rev. Genet. 2007; 8:857–868. [PubMed: 17943193]
39. Bustamante CD, et al. Natural selection on protein-coding genes in the human genome. Nature. 2005; 437:1153–1157. [PubMed: 16237444]
40. Blekhman R, et al. Natural selection on genes that underlie human disease susceptibility. Curr. Biol. 2008; 18:883–889. [PubMed: 18571414]
41. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. Natural selection has driven population differentiation in modern humans. Nat. Genet. 2008; 40:340–345. [PubMed: 18246066]
42. Akey JM, et al. Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol. 2004; 2:e286. [PubMed: 15361935]
43. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat. Rev. Genet. 2010; 11:17–30. [PubMed: 19953080]
44. Vasseur E, Quintana-Murci L. The impact of natural selection on health and disease: uses of the population genetics approach in humans. Evol. Appl. 2013; 6:596–607. [PubMed: 23789027]
45. Chiaroni J, Underhill PA, Cavalli-Sforza LL. Y chromosome diversity, human expansion, drift, and cultural evolution. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:20174–20179. [PubMed: 19920170]
46. Baten J, Blum M. Growing Tall but Unequal: New Findings and New Background Evidence on Anthropometric Welfare in 156 Countries, 1810–1989. Econ. Hist. Dev. Reg. 2012; 27:S66–S85.
47. Ovaskainen 0, Karhunen M, Zheng C, Arias JMC, Merila J. A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. Genetics. 2011; 189:621–632. [PubMed: 21840853]

48. Diverse populations collaborative. Weight-height relationships and body mass index: some observations from the Diverse Populations Collaboration. Am. J. Phys. Anthropol. 2005; 128:220–229. [PubMed: 15761809]
49. Lande R. Genetic Variation and Phenotypic Evolution During Allopatric Speciation. Am. Nat. 1980:463–479.
50. Esko T, et al. Genetic characterization of northeastern Italian population isolates in the context of broader European genetic diversity. Eur. J. Hum. Genet. 2013; 21:659–665. [PubMed: 23249956]
51. Weir BS, Hill WG. Estimating F-statistics. Annu. Rev. Genet. 2002; 36:721–750. [PubMed: 12359738]
52. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. 2008
53. Cockerham CC, Weir BS. Correlations, descent measures: drift with migration and mutation. Proc. Natl. Acad. Sci. 1987; 84:8512–8514. [PubMed: 3479805]
54. Karhunen M, Ovaskainen 0, Herczeg G, Merila J. Bringing habitat information into statistical tests of local adaptation in quantitative traits: a case study of nine-spined sticklebacks. Evolution. 2014; 68:559–568. [PubMed: 24117061]
55. Abercasis GR, Cardon LR, Cookson WOC. A general test of association for quantitative traits in nuclear families. Am J. Hum Genet. 2000; 66:279–292. [PubMed: 10631157]
56. 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. 2010; 107(Suppl):1752–1756. [PubMed: 20133822]
57. Elhaik E. Empirical distributions of F(ST) from large-scale human polymorphism data. PLoS One. 2012; 7:e49837. [PubMed: 23185452]
58. Jakobsson M, Edge MD, Rosenberg NA. The relationship between F(ST) and the frequency of the most frequent allele. Genetics. 2013; 193:515–528. [PubMed: 23172852]
59. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr. Biol. 2010; 20:R208–R215. [PubMed: 20178769]
60. Williams AL, Patterson N, Glessner J, Hakonarson H, Reich D. Phasing of many thousands of genotyped samples. Am. J. Hum. Genet. 2012; 91:238–251. [PubMed: 22883141]
61. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda). 2011; 1:457–470. [PubMed: 22384356]
62. Abecasis GR, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]
63. Delaneau 0, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. Nature Methods. 2012; 9:179–181. [PubMed: 22138821]
64. Hadfield, JD. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. Stat. Softw. 2010. at <http://www.jstatsoft.org/v33/i02/paper>

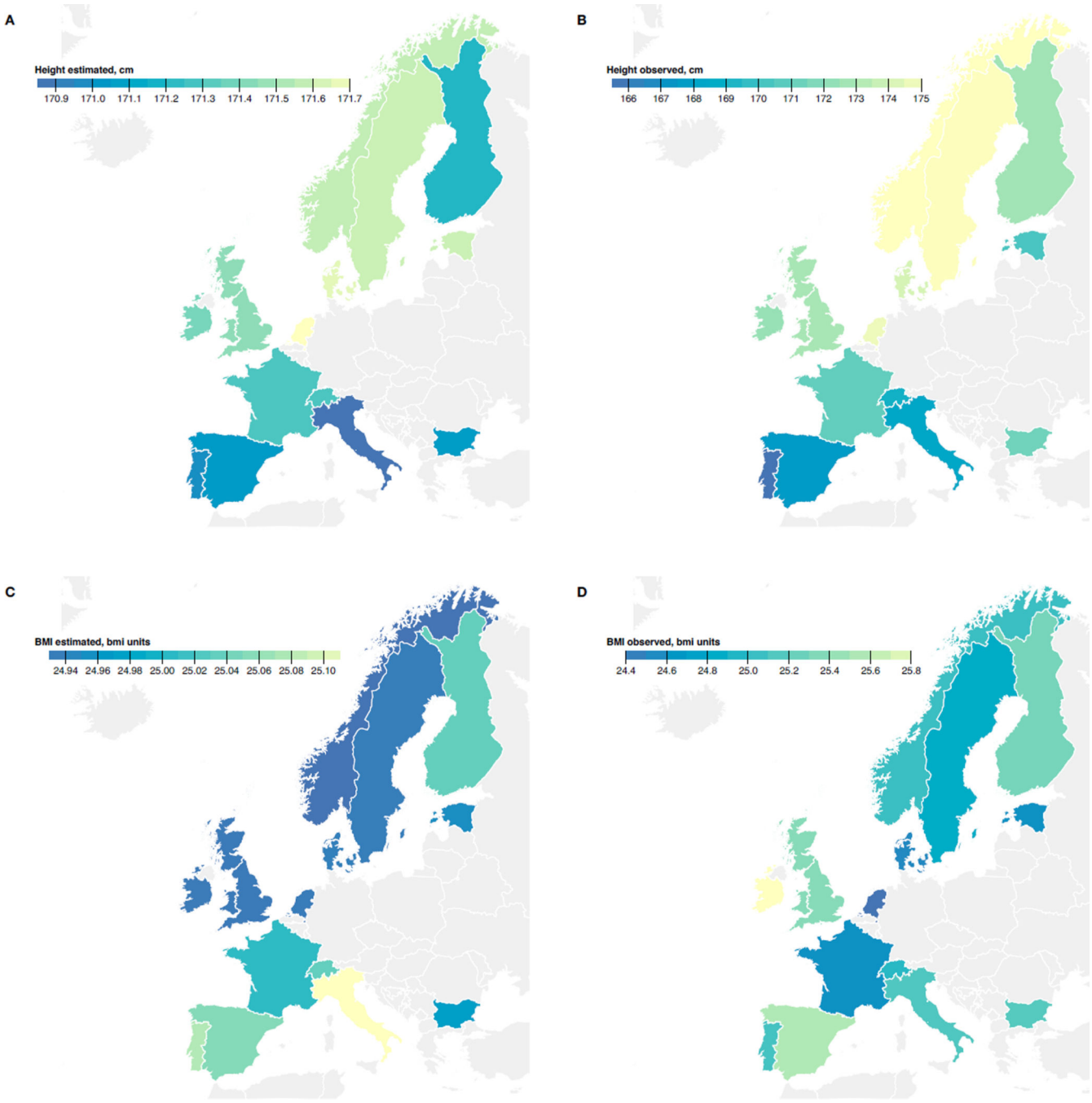
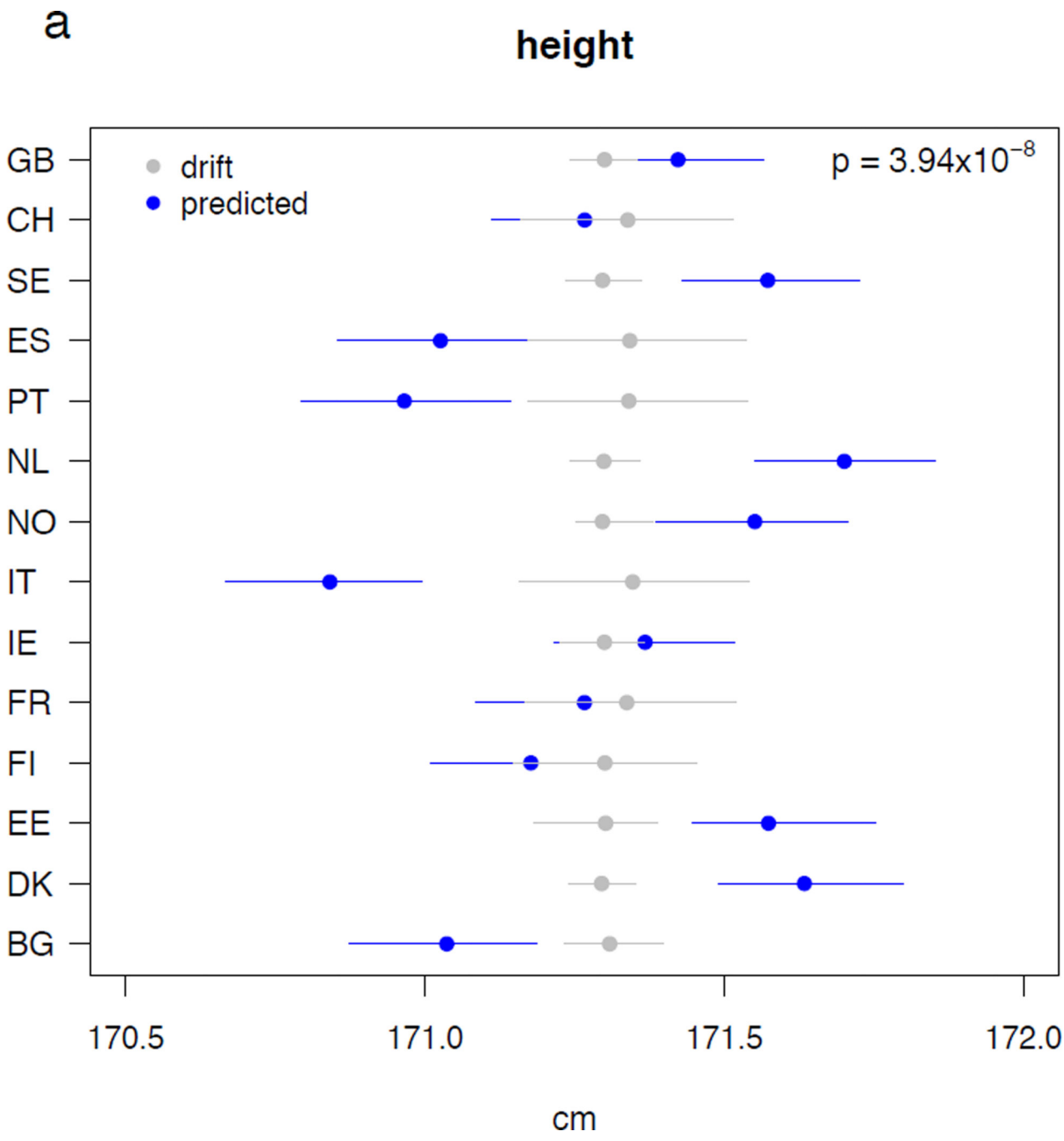
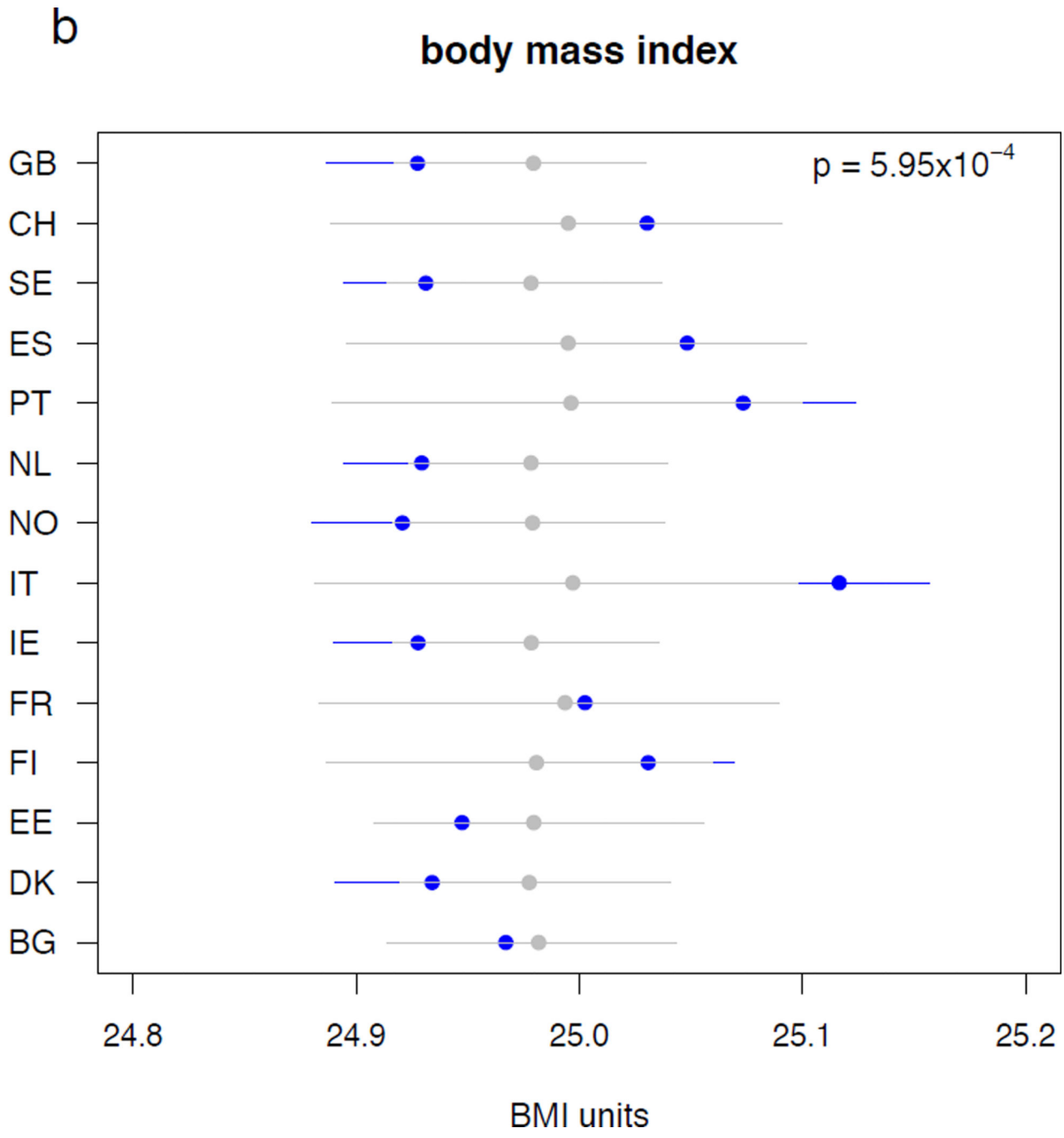


Figure 1. Observed divergence and predicted genetic divergence in height and body mass index (BMI) and height across 14 European nations

Across Europe, the observed means and predicted genetic means for height and BMI of 14 European nations are shown. From recently published data, we estimated national differences in mean height and BMI for 14 European nations accounting for time trends (Figure 1), with a European average height of 171.1 (95% CI: 169.6, 172.8) and average BMI of 25.0 (95% CI: 24.7, 25.3) across nations for males between the years 2000 and 2010.





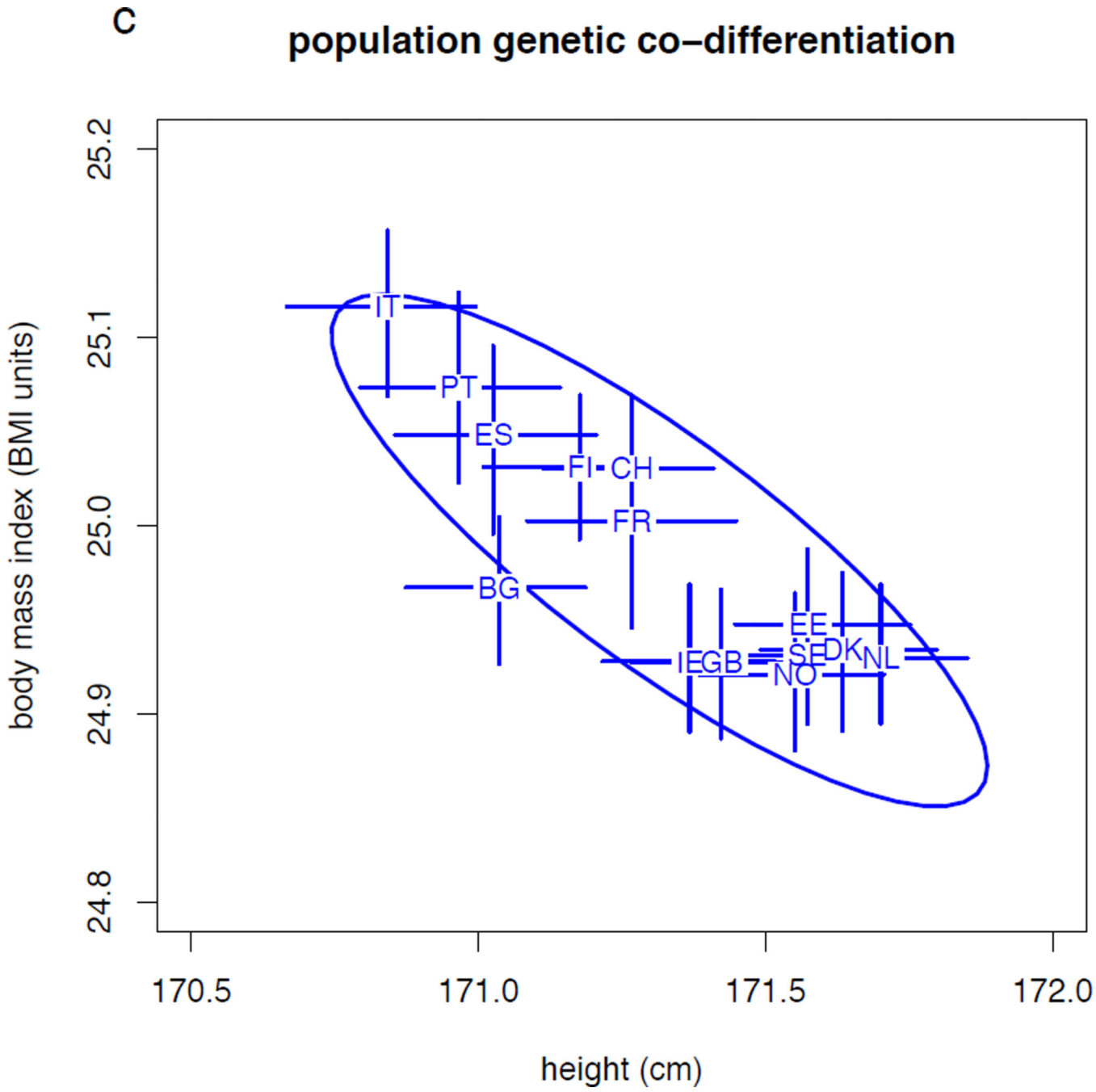
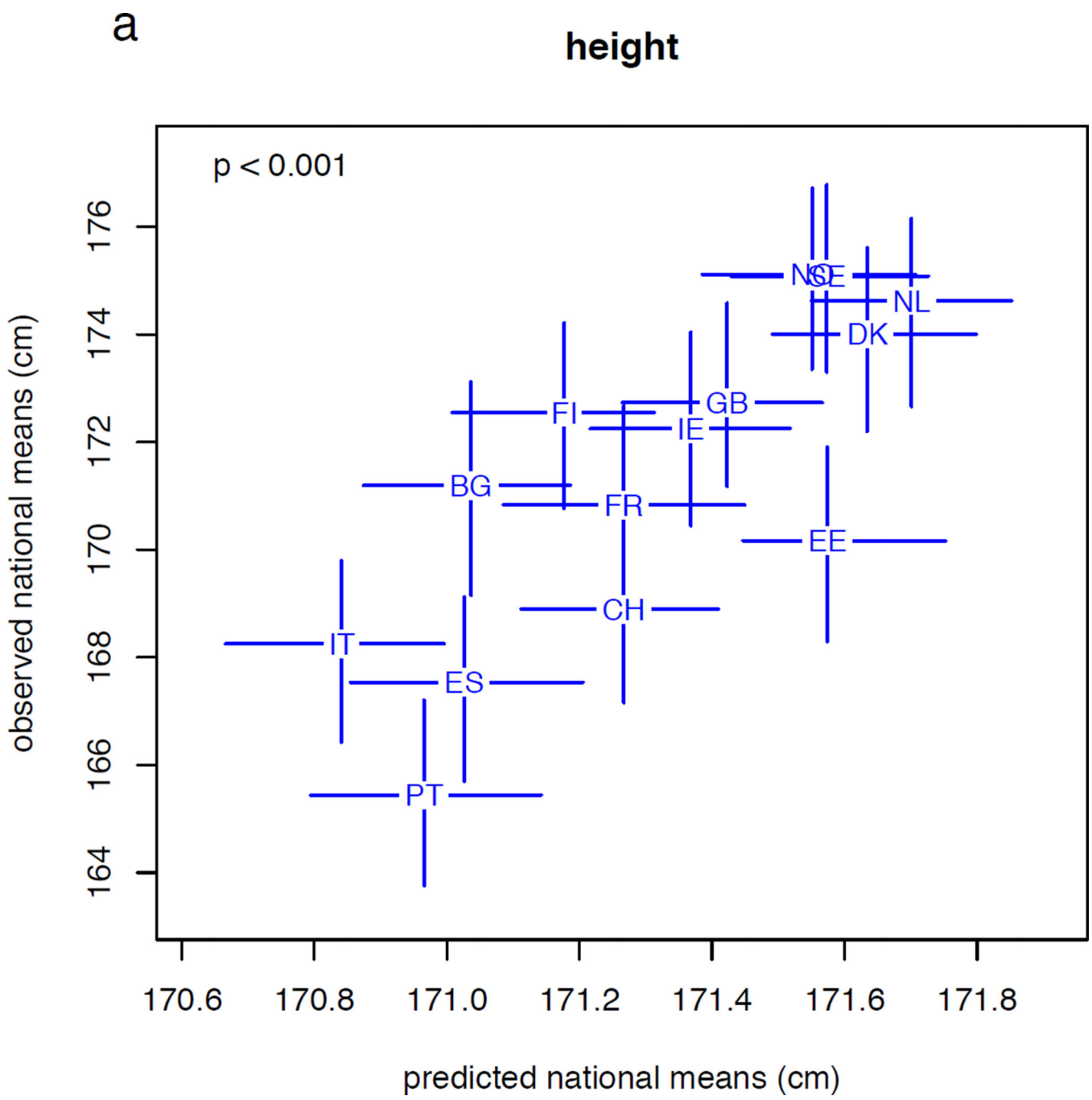


Figure 2. Predicted genetic differentiation compared to the expectation under drift for height and body mass index across 14 European nations

Mean predicted genetic (blue) and null model (grey) values of 14 European nations are shown, with 95% credible intervals, for (a) height (cm) and (b) body mass index (BMI units). ISO2 country codes indicate each nation. The average p-value of differentiation from the null expectation is $p<4.3e^{-14}$ for height and $p<8.7e^{-07}$ for BMI. (c) Pattern of population co-differentiation of height and body mass index across 14 European nations (blue). The negative population genetic co-differentiation of −0.80 (95% CI: −0.95, −0.60) is represented by a blue ellipse.



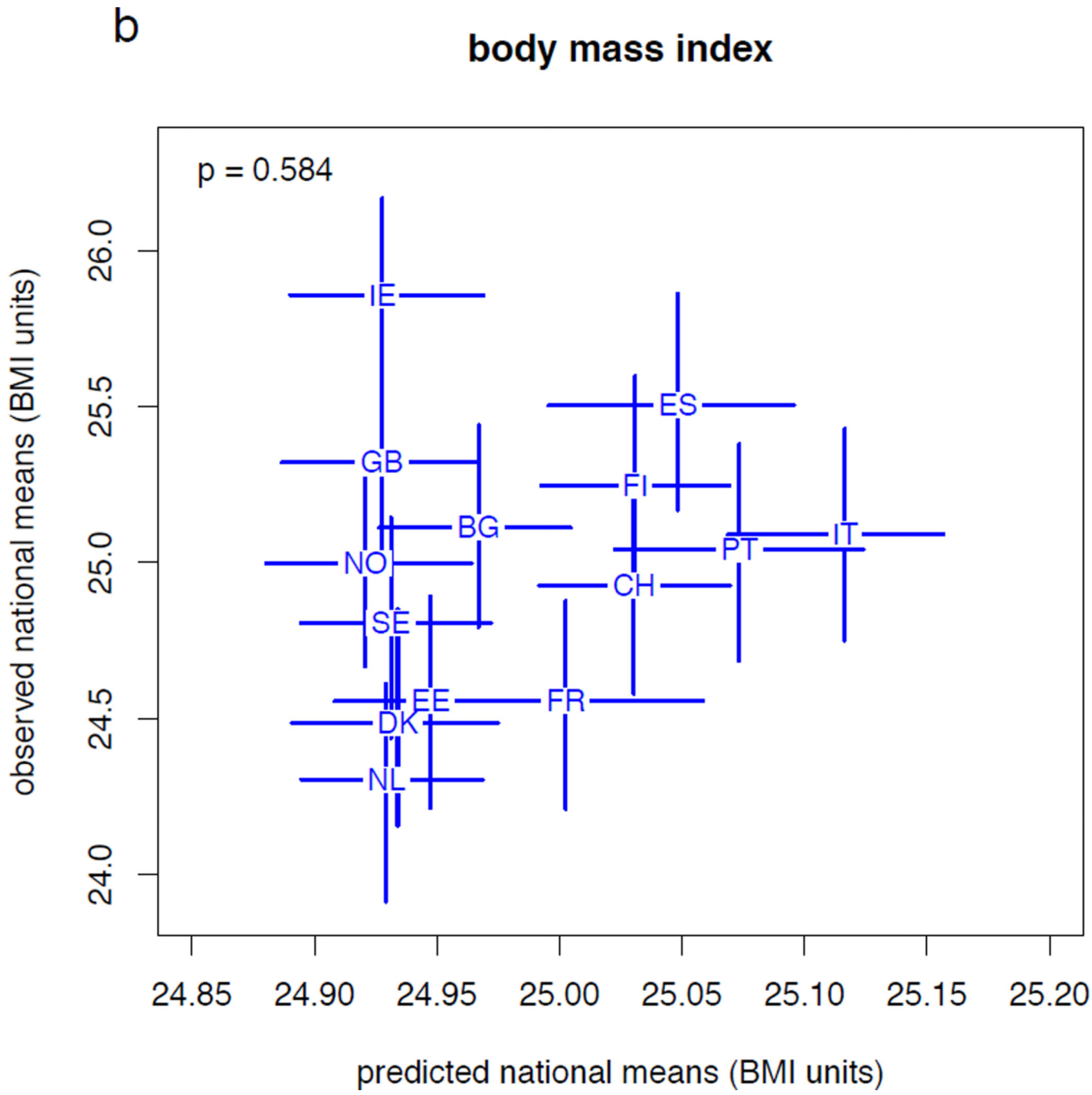


Figure 3. Association between observed population means and predicted genetic population means for height and body mass index across 14 European nations

Predicted population genetic means and observed population means for (a) height and (b) body mass index (BMI). *P* values give the significance of the multivariate Pearson product moment correlation between the predicted population genetic means and the observed population means for both traits. For height, the correlation (r = 0.51; 95% CI 0.39, 0.61) was greater than that expected under the null model (r = 0.03, 95% CI −0.21, 0.17). For

BMI, the correlation (r = −0.10, 95% CI −0.19, 0.01) was not significantly different from the null expectation (r = −0.08, 95% CI −0.24, 0.15).