Population genetic differentiation of height and body mass index across Europe

Matthew R Robinson¹, Gibran Hemani¹, Carolina Medina-Gomez², Massimo Mezzavilla³,⁴, Tonu Esko⁵–⁸, Konstantin Shakhbazov¹, Joseph E Powell¹,⁹, Anna Vinkhuyzen¹, Sonja I Berndt¹⁰, Stefan Gustafsson¹¹, Anne E Justice¹², Bratati Kahali¹³,¹⁴, Adam E Locke¹⁵, Tune H Pers⁶–⁸,¹⁶, Sailaja Vedantam⁶,⁷, Andrew R Wood¹⁷, Wouter van Rheenen¹⁸, Ole A Andreassen¹⁹, Paolo Gasparini³,⁴, Andres Metspalu⁵, Leonard H van den Berg¹⁸, Jan H Veldink¹⁸, Fernando Rivadeneira², Thomas M Werge²⁰–²², Goncalo R Abecasis¹⁵, Dorret I Boomsma²³–²⁵, Daniel I Chasman²⁶, Eco J C de Geus²³–²⁵, Timothy M Frayling¹⁷, Joel N Hirschhorn⁵–⁸, Jouke Jan Hottenga²³–²⁵, Erik Ingelsson¹¹,²⁷, Ruth J F Loos²⁸–³¹, Patrik K E Magnusson³², Nicholas G Martin³³, Grant W Montgomery³³, Kari E North¹³,¹⁴,³⁴, Nancy L Pedersen³², Stefan Gustafsson¹¹, Konstantin Shakhbazov¹, Joseph E Powell¹,⁹, Anna Vinkhuyzen¹, Sonja I Berndt¹⁰, Stefan Gustafsson¹¹, Anne E Justice¹², Bratati Kahali¹³,¹⁴, Adam E Locke¹⁵, Tune H Pers⁶–⁸,¹⁶, Sailaja Vedantam⁶,⁷, Andrew R Wood¹⁷, Wouter van Rheenen¹⁸, Ole A Andreassen¹⁹, Paolo Gasparini³,⁴, Andres Metspalu⁵, Leonard H van den Berg¹⁸, Jan H Veldink¹⁸, Fernando Rivadeneira², Thomas M Werge²⁰–²², Goncalo R Abecasis¹⁵, Dorret I Boomsma²³–²⁵, Daniel I Chasman²⁶, Eco J C de Geus²³–²⁵, Timothy M Frayling¹⁷, Joel N Hirschhorn⁵–⁸, Jouke Jan Hottenga²³–²⁵, Erik Ingelsson¹¹,²⁷, Ruth J F Loos²⁸–³¹, Patrik K E Magnusson³², Nicholas G Martin³³, Grant W Montgomery³³, Kari E North¹³,¹⁴,³⁴, Nancy L Pedersen³², Timothy D Spector³⁵, Elizabeth K Speliotes¹⁵, Michael E Goddard³⁶,³⁷, Jian Yang¹,⁹ & Peter M Visscher¹,⁹

Across-nation differences in the mean values for complex traits are common¹–⁸, but the reasons for these differences are unknown. Here we find that many independent loci contribute to population genetic differences in height and body mass index (BMI) in 9,416 individuals across 14 European countries. Using discovery data on over 250,000 individuals and unbiased effect size estimates from 17,500 sibling pairs, we estimate that 24% (95% credible interval (CI) = 9%, 41%) and 8% (95% CI = 4%, 16%) of the captured additive genetic variance for height and BMI, respectively, reflect population genetic differences. Population genetic divergence differed significantly from that in a null model (height, \(P < 3.94 \times 10^{-8}\); BMI, \(P < 5.95 \times 10^{-4}\)), and we find an among-population genetic correlation for tall and slender individuals (\(r = -0.80, 95\% \text{ CI} = -0.95, -0.60\)), consistent with correlated selection for both phenotypes. Observed differences in height among populations reflected the predicted genetic means (\(P = 0.51; P < 0.001\)), but environmental differences across Europe masked genetic differentiation for BMI (\(P < 0.58\)).

Many of the phenotypes that vary within human populations are complex, in that they are determined by alleles at multiple loci and many non-genetic factors⁹–¹⁵. Therefore, it is reasonable to assume that regional differences in such traits have a complex basis¹⁶–¹⁸. Understanding these regional differences requires knowledge of the relative roles of environmental and 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 are explained by population genetic effects¹⁹. Thus far, these estimates have yet to be made outside of laboratory study populations²⁰, and experimental designs for human populations have been lacking because of confounding from genetic and environmental effects.

At least 135 million European citizens are obese²¹, resulting in major direct and indirect health and economic costs²⁶,²²,²³. Regional differences across Europe in height and susceptibility to weight gain, as defined by BMI, are well documented⁵,¹⁸,²²–²⁶. Regional differences across Europe in height and susceptibility to weight gain, as defined by BMI, are well documented⁵,¹⁸,²²–²⁶. Regional differences across Europe in height and susceptibility to weight gain, as defined by BMI, are well documented⁵,¹⁸,²²–²⁶. Regional differences across Europe in height and susceptibility to weight gain, as defined by BMI, are well documented⁵,¹⁸,²²–²⁶.

For height, there is empirical evidence for selection at height-associated loci, but the extent to which common loci contribute to population genetic variance or the observed regional phenotypic differences remains unknown (although see ref. 26).

Here we estimate cumulative population genetic differentiation for height and BMI captured by multiple unlinked loci across 9,416 Europeans from 14 countries, using population genetic analyses (Online Methods and Supplementary Figs. 1 and 2). We performed genome-wide association study (GWAS) meta-analyses on data from recent studies, selecting independent loci (linkage disequilibrium (LD) \(D^2 <0.1\) and >1 Mb apart³⁵) associated with either trait in a European-ancestry sample (~250,000 individuals for height and ~350,000 individuals for BMI). We reestimated the effects of each SNP in a within-family study design, which is unbiased by population stratification, and used these effect sizes to create a genetic predictor for both phenotypes (also termed a ‘profile’ or ‘polygenic score’)³⁶.
The proportion of variance in the genetic predictor attributable to population differences was estimated in a Bayesian mixed-effects model, alongside the co-differentiation of the phenotypes and the predicted means for each nation (Online Methods). Using theory and simulation study, we show that, whereas within-family effect sizes are unbiased, population genetic analyses conducted using loci ascertained from a standard GWAS can be biased if population stratification is not fully accounted for (Online Methods, Supplementary Figs. 3–5 and Supplementary Note). There is no certainty that population stratification is completely controlled for in large-scale meta-analyses, and we thus repeated our analysis using (i) non-ascertained, unlinked (LD $r^2 < 0.1$ and >1 Mb apart), common (minor allele frequency (MAF) >1%) HapMap 3 loci (~40,000 SNPs) and (ii) unlinked (LD $r^2 < 0.1$ and >1 Mb apart), common (MAF >1%) HapMap 3 loci selected on the basis of their within-family association with each phenotype (~40,000 SNPs for both traits). This analysis provides an unbiased, genome-wide estimate, representing a lower limit of population genetic differentiation at common, unlinked loci.

The maximum proportion of variance in a polygenic predictor attributable to population genetic differences was 24% (95% CI = 9%, 41%) and 8% (95% CI = 4%, 16%) for height and BMI, respectively, 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 low $P$ value in the meta-analysis (Supplementary Fig. 6). For BMI, the continual addition of SNPs increased the proportion of population-level variance captured (Supplementary Fig. 6). For both traits, the among-population variation was greater in predictors that explained greater phenotypic variance (Supplementary Fig. 7). Our results were confirmed using the non-ascertained independent, genome-wide loci (height: 8.6%, 95% CI = 3%, 15.7%; BMI: 8.8%, 95% CI = 1.1%, 5.3%) and the set of independent, genome-wide loci selected on the basis of their within-family association (height: 11.9%, 95% CI = 4.5%, 21.8%; BMI: 8%, 95% CI = 3.4%, 14.7%). The lower among-population variance captured using non-ascertained loci reflects reduced prediction accuracy, likely due to the addition of a large number of loci with no detectable association. Subsequent results are presented using the predictor for each trait that captured the greatest amount of population- and individual-level variance (comprising 2,660 SNPs for height and 11,919 SNPs for BMI); however, the results remained the same irrespective of the SNPs selected (Supplementary Fig. 8). The predicted population genetic means for the traits are shown in Figure 1 alongside the observed values, estimated from an independent set of recently published data, accounting for trends over time. Among-population differences in allele frequency are expected to create genetic differences in height, such that people from the Netherlands are on average 1 cm taller than those from Italy, and genetic differences in BMI, such that, on average, the BMIs for people from Italy and Denmark differ by 0.2 units (Figs. 1 and 2).

Genetic differences among populations may occur by random, chance processes or through natural selection in the evolutionary past. We thus compared our estimates to the values from a null quantitative genetic model of multivariate population differentiation. We found strong evidence that the divergence of each trait was greater than expected under the neutral model (Fig. 2). The overall level of neutral genetic differentiation was small for both height (1.2%; 95% CI = 0.01%, 1.78%) and BMI (1.9%, 95% CI = 0.48%, 2.97), reflecting the average $F_{ST}$ (a measure of population differentiation due to genetic structure; Supplementary Note) of the SNP sets between the populations of 1% for height and 1.2% for BMI. Our results were confirmed using non-ascertained independent, genome-wide loci (height, $P = 3.29 \times 10^{-8}$; BMI, $P = 0.018$) and independent, genome-wide loci selected on the basis of their within-family association (height, $P = 2.67 \times 10^{-8}$; BMI, $P = 8.35 \times 10^{-5}$). We therefore reject the null model, and our results suggest that population genetic differentiation across these 14 European countries for height and BMI has been driven by selection on standing genetic variation across geographical regions in the 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...
For BMI, 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). These results imply that selection has acted on common loci to increase height while reducing BMI and vice versa, and a genetic predisposition for tall stature at the population level was associated with a genetic predisposition for slenderness (low BMI). As height and BMI are nearly uncorrelated at the individual level (correlation among genetic profile scores within populations r = −0.016, 95% CI = −0.041, 0.001), selection for one trait should not elicit a response in the other. Our results suggest that selection has acted on both phenotypes, although, as some genes affect both phenotypes, we cannot rule out differentiation in one trait having been mediated by selection for the other. The population genetic co-divergence shown here is inconsistent with random genetic drift because the expectation with drift is that the among-population genetic correlation will equal the within-population correlation.47,49

We tested whether the observed phenotypic differences across the 14 European countries reflect genetic differentiation at common loci or whether current environmental differences among countries (in diet, economy, climate, etc.) mask the population genetic differences.
differences 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 (Fig. 3), suggesting that the phenotypic differences that we observe across countries reflect differences in allele frequency at common height-associated loci. For BMI, the observed pattern did not reflect the pattern of population genetic differentiation \( (r = -0.10, 95\%\ CI = -0.19, 0.01; P = 0.584; \text{Fig. 3}) \). This suggests that, although selection has created population genetic differentiation for BMI, environmental differences among countries mask this population genetic differentiation. We found no evidence of an association between the genetic differentiation expected under drift and the observed values (Fig. 3), implying that the observed national patterns do not reflect genetic drift.

We identified the loci that contributed the most to the population genetic differentiation for both traits. We found that the SNPs making the largest contributions to the phenotypic variance were enriched for association with the genome-wide pattern of population genetic differentiation (Supplementary Figs. 9 and 10), a pattern expected on the basis of our theory (Supplementary Figs. 11 and 12) and supported by the fact that the proportion of population-level variance was greater in a predictor explaining greater phenotypic variance. We found no evidence for significant association of population differentiation at any SNP across the genome, implying that many loci of small effect combine across the genome to create the detected genome-wide population genetic differentiation. Annotation of the 500 SNPs contributing the most to each trait with genes suggests that population genetic variation across the 14 countries for height and BMI is underlain by the combined effects of multiple pathways, with an overlap in the genes involved (Supplementary Fig. 13).

Finally, we examined population genetic differentiation in height and BMI on both a local and worldwide scale. We found no evidence for population genetic differentiation across six northern Italian villages\(^{39} \) (Supplementary Fig. 14). We then examined population differentiation for height and BMI in the Human Genetic Diversity Panel (HGDP), as used in a previous study\(^{32} \) (Supplementary Fig. 15). For both phenotypes, we found evidence to reject the null hypothesis that population genetic differentiation reflects neutrality (Supplementary Fig. 15), extending previous work that reported no significant differentiation for BMI using a limited number of loci\(^{32} \). Additionally, we found no evidence for population genetic co-differentiation of height and BMI, implying a European-specific pattern of selection.

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) population genetic differentiation for height and BMI does not reflect a pattern expected under neutrality, and selection has thus driven the differences observed; (iii) population genetic divergences for height and BMI are correlated within Europe, with this correlation reaching a greater level than expected under neutrality, implying that the height-associated loci under selection are enriched for loci with effects that reduce BMI; and (iv) selection-driven population genetic differences for height reflect the phenotypic patterns we see across Europe, whereas, for BMI, environmental factors are masking the population genetic differences. Although genotype-environment effects and rare variants will have a role in shaping population genetic differentiation, the focus of our 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 buildups for any set of phenotypes.

**METHODS**

Methods and any associated references are available in the online version of the paper.

---

**ACKNOWLEDGMENTS**

We thank the reviewers for their very helpful and insightful comments that greatly improved the manuscript. We also thank B. Hill and O. 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 DE13100691. J.Y. is supported by a Charles and Sylvia Viertel Senior Medical Research Fellowship and by NHMRC grant 1052864. 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 set, which includes data from many cohorts and received support from many funding bodies. TWINGE was supported by the Swedish Research Council (M-2003-1112), GenomEUwin (EU/QLRT-2001-01254 and QLG2-CT-2002-01254), US National Institutes of Health (NIH) grant DK U01-066134, the Swedish Foundation for Strategic Research (SSF), and the Heart and Lung Foundation (20070481). For the Netherlands Twin Register (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 and Spinozapremie 56-464-14192), the Center for Medical Systems Biology (CSMB; WGO 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 (ESP; EU/QLRT-2001-01254), the European Community’s Seventh Framework Programme (FP7/2007-2013) under the ENGAGE project grant agreement (HEALTH-F4-2007-201413), the European Research Council (ERC Advanced; 230374), the Rutgers University Cell and DNA Repository (National Institute for Mental Health (NIMH)), U24-MH068457-06, the Avera Institute (Sioux Falls, South Dakota, USA) and the US NIH (ROI-D0042157-01A, 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 and the European Community’s Seventh Framework Programme (FP7/2007-2013) under the ENGAGE project grant agreement (HEALTH-F4-2007-201413). TwinsUK also receives support from the UK Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy’s and St Thomas’ National Health Service (NHS) Foundation Trust in partnership with King’s College London. T.D.S. is the holder of an ERC Advanced Principal Investigator award. Genotyping for the TwinsUK study was performed by the Wellcome Trust Sanger Institute, with the support of the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre award to the Wellcome Trust (grant agreement 1048853 and 1050218). J.E.P. is supported by Australian Research Council grant DE13100691. J.Y. is supported by a Charles and Sylvia Viertel Senior Medical Research Fellowship and by NHMRC grant 1052864. 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 set, which includes data from many cohorts and received support from many funding bodies. TWINGE was supported by the Swedish Research Council (M-2003-1112), GenomEUwin (EU/QLRT-2001-01254 and QLG2-CT-2002-01254), US National Institutes of Health (NIH) grant DK U01-066134, the Swedish Foundation for Strategic Research (SSF), and the Heart and Lung Foundation (20070481). For the Netherlands Twin Register (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 and Spinozapremie 56-464-14192), the Center for Medical Systems Biology (CSMB; WGO 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 (ESP; EU/QLRT-2001-01254), the European Community’s Seventh Framework Programme (FP7/2007-2013) under the ENGAGE project grant agreement (HEALTH-F4-2007-201413), the European Research Council (ERC Advanced; 230374), the Rutgers University Cell and DNA Repository (National Institute for Mental Health (NIMH)), U24-MH068457-06, the Avera Institute (Sioux Falls, South Dakota, USA) and the US NIH (ROI-D0042157-01A, 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 and the European Community’s Seventh Framework Programme (FP7/2007-2013) under the ENGAGE project grant agreement (HEALTH-F4-2007-201413). TwinsUK also receives support from the UK Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy’s and St Thomas’ National Health Service (NHS) Foundation Trust in partnership with King’s College London. T.D.S. is the holder of an ERC Advanced Principal Investigator award. Genotyping for the TwinsUK study was performed by the Wellcome Trust Sanger Institute, with the support of the National Eye Institute via a US NIH/Center for Inherited Disease Research (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. Funding for SHARE Illumina genotyping was provided under an agreement between Illumina and Boston University. The QIMR researchers acknowledge funding from the Australian NHMRC (grants 241944, 389875, 389891, 389902, 389928, 442915, 442981, 496739, 496888 and 552485) and the US NIH (grants AA07535, AA10248, AA014041, AA13320, AA13321, AA13326 and...
ONLINE METHODS

Genetic differentiation of height and BMI across Europe. The theoretical basis of our analysis framework builds upon those from refs. 32,47,51–53 and is outlined in the Supplementary Note. We apply this framework to examine population genetic differentiation in height and BMI across Europe. An overview of the analysis steps is provided in Supplementary Figure 1.

Imputation. All of the cohorts used in this study were independently imputed to the 1000 Genomes Project reference panel, using identical quality control procedures on the initial data sets of per-SNP missing data rate <0.02, MAF >0.01, per-individual missing data rate <0.05 and Hardy-Weinberg disequilibrium P <0.0001. Imputation for the majority of the cohorts was performed in two stages. First, the target data were assigned to haplotypes using HAPI-UR54. Second, IMPUTE2 (ref. 55) was used to impute the haplotypes to the 1000 Genomes Project reference panel56 (release 1, version 3). We then selected SNPs that were present across all data sets at an imputation information score of >0.8. The imputation for the Netherlands cohort was identical, except that SHAPEIT2 (ref. 57) was used for haplotyping. We performed these same quality control steps again after combining the data from the 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 studies33,34 to select independent loci (r² <0.1 and >1 Mb apart using the PLINK clumping procedure35) that were associated with the traits in a large sample of individuals (~250,000 for height and ~350,000 for BMI) of European ancestry. We excluded cohorts overlapping with our within-family and prediction samples. For height, we excluded the TWINGENE study, the TwinsUK study, the QIMR sample, the Framingham Heart Study sample and the Netherlands Twin Register. For BMI, cohort-level summary statistics were not available for all samples, and we could only exclude the QIMR sample.

Within-family estimation of SNP effects. We reestimated the SNP effects at these loci in a within-family sibling pair data set (Supplementary Table 1) using the QFAM procedure in PLINK described in equation (3.1) in the Supplementary Note.

Population genetic analyses conducted using ascertained loci from a standard GWAS can be biased if population stratification is not fully accounted for (Supplementary Note). Thus, to confirm our results, we also selected a non-ascertained genome-wide set of unlinked (LD r² <0.1 and >1 Mb apart), common (MAF >1%) HapMap 3 loci for height and BMI (i.e., ~40,000 SNP loci) that passed quality control in both the within-family and prediction samples. We estimated the effects of these SNPs again using our within-family sibling pair data set with the QFAM procedure in PLINK described in equation (3.1) in the Supplementary Note. Additionally, we used the clumping procedure in PLINK to select genome-wide unlinked (LD r² <0.1 and >1 Mb apart), common (MAF >1%) HapMap 3 loci on the basis of their within-family association with height and BMI (~40,000 SNP loci for both traits).

Genomic profile scoring in a collection of European genomic data. We used the within-family effect sizes to create genetic predictors for individuals across a collection of European genomics data. All data were 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 sampled individuals. From the Estonian and Finnish cohorts, we selected 1,000 individuals at random who were included in all analyses. For the Netherlands cohort, we selected 1,000 control individuals from the MiNE ALS study. 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 Wellcome Trust Case Control Consortium 2, we used control individuals from Bulgaria, Ireland, Norway, Denmark, Sweden and the UK (Supplementary Table 1). The POPRES data set used in this manuscript was obtained from the database of Genotypes and Phenotypes (dbGaP) through accession phs000145.v4.p2.

We used the genetic predictors for height and BMI as response variables in a bivariate Bayesian mixed-effects model, which is outlined in equation (2.2) of the Supplementary Note. This model was estimated using the R package MCMCglmm58, with uninformative inverse Wishart priors, a burn-in period of 7,000 iterations, a sampling interval of 10 iterations and a total of 17,000 iterations, providing 1,000 posterior estimates. This provided estimates of the population genetic (co)variance of height and BMI, the residual (individual-level) (co)variance and the best linear unbiased predictors of the means of each nation along with the 95% credible intervals. For each set of SNPs, the genetic predictor was standardized to a z score to allow comparison across SNPs and comparison with the null model.

Comparison with a null model. At each stage of the analysis, the estimates obtained were compared to the expectations under random genetic drift, using a quantitative genetic framework for studying population differentiation that is outlined in the Supplementary Note.

In brief, the within-family regression coefficients were randomized across SNPs 1,000 times, and 1,000 genetic predictors were created in the European prediction sample. By keeping the effect sizes consistent but attributing these effects across SNPs at random, the genetic predictors generated reflect the action of genetic drift. Second, each set of genetic predictors was standardized to a z score and used as a response variable in the Bayesian mixed-effects model outlined in equation (2.2) of the Supplementary Note. This provided 1,000 estimates of the population genetic variance and population means under drift; these values are displayed in the figures as the estimates from the neutral model. Third, the sample covariance matrix of these 1,000 estimates was calculated, which provided an estimate of the expected population-level covariance in phenotype under drift. We then used a Mahalanobis distance statistic to provide a measure of the relative deviation of our predicted population-level means from their multivariate theoretical expectations under drift (Supplementary Note). This calculation provided the χ² test statistic used to compare our predicted estimates to the expected values under drift. As both the drift profile and trait profile scores were transformed to a z score, this comparison was on the same standard deviation scale.

European phenotypic data. Height and BMI measures for males for each of the 14 European countries were taken from recently published estimates26,37. For height, measures were available from 1860 to the present day, and, for BMI, measures were available from 1980 to the present day. Values for both phenotypes were adjusted for trends over time before estimating the population means within a mixed-effects model. The model-generated estimates of these means were then compared to the predicted genetic means as described in equation (5.1) to equation (5.4) in the Supplementary Note.

Transforming population means onto the observed scale. For graphical presentation of the results, each genetic predictor q, created for each individual i for each trait m, was approximately transformed back to the observed scale as outlined in equation (6.1) in the Supplementary Note. We used an independent sample of population data to determine the amount of phenotypic variance explained by the genetic predictors created from different sets of SNPs, as measured by cov(zq, zm), where zq and zm are the z scores of the phenotype and genetic predictor, respectively, in equation (6.1) of the Supplementary Note. Individuals within the Health and Retirement Study (HRS) (Supplementary Table 1) were unrelated, and phenotypic values were 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 standard deviation for height and BMI measured in the sample of 17,500 quasi-independent sibling pairs was estimated accounting for sex differences. The HRS data set was obtained from dbGaP through accession phs000428.v1.p1.

Testing the contribution of each SNP to the pattern of population differentiation. As described in equation (4.4) in the Supplementary Note, we estimated a χ² value for each SNP representing its contribution to the pattern of population genetic differentiation. We tested for the association between a SNP’s contribution to differentiation and its MAF and expected contribution to phenotypic variation estimated as 2pqβ², where p is the frequency of the minor allele, q is the frequency of the major allele and β is the within-family effect size.
Positional and functional annotation of SNPs. Our aim was to describe the positional genetic annotation and gene ontology (GO) categories for the height- and BMI-associated SNPs that contributed the most to the genome-wide pattern of population genetic variance. We selected the 500 height- and BMI-associated SNPs that contributed the most to differentiation. First, on the basis of the genomic position of these SNPs, we assigned them to genes and simply estimated the number of overlapping genes involved in the genetic differentiation of height and BMI. Second, SNPs were assigned to the following positional genetic annotation categories: 3′ UTR variant, 5′ UTR variant, intronic variant, noncoding transcription variant, variant 1–1,000 bp downstream of a gene, variant 1–1,000 bp upstream of a gene, missense transcription variant, synonymous transcription variant and noncoding exonic variant. Finally, SNPs were assigned Ensemble gene identities and then GO terms. The top GO annotations are listed in Supplementary Table 2.

We then conducted statistical testing. As a baseline, we used the top 10,000 SNPs for height and BMI. We first compared the number of SNPs among the 500 SNPs contributing the most to differentiation in each genic category, to count data from the top 10,000 SNPs. This provided a list of genetic categories potentially enriched for highly differentiated SNPs. We used Fisher’s exact tests (hypergeometric) with Bonferroni P-value correction. Second, we compared the number of SNPs among the 500 SNPs contributing the most to differentiation in each GO term, to count data from the top 10,000 SNPs. Again, we used Fisher’s exact tests with Bonferroni P-value correction. From this analysis, we selected a list of the top 20 functional categories potentially enriched for each trait. We assigned P values to our comparisons of count data but used these only as a guide to select the top categories, rather than as a definitive test of enrichment. All annotation was conducted using the R library biomaRt from Bioconductor.

Genomic profile scoring worldwide. We repeated our analyses for height and BMI using data from the HGDP as analyzed in ref. 32. We imputed the data following our protocol outlined above.

Simulation study using real genotype data. In addition to the simulations described in the Supplementary Note, we used the common, independent HapMap 3 SNPs from the 17,500 sibling pairs in the main empirical analyses (Supplementary Table 1) as the basis for a series of simulations. We randomly assigned 5,000 of the independent loci across the genome to be causal variants, with their effects sampled from a normal distribution with a mean of 0 and a variance of 1. The heritability of the trait was simulated to be 90%, described in the simulation study using real genotype data. We used Fisher’s exact tests (hypergeometric) with Bonferroni P-value correction. Second, we compared the number of SNPs among the 500 SNPs contributing the most to differentiation in each genic category, to count data from the top 10,000 SNPs. This provided a list of genetic categories potentially enriched for highly differentiated SNPs. We used Fisher’s exact tests (hypergeometric) with Bonferroni P-value correction. Second, we compared the number of SNPs among the 500 SNPs contributing the most to differentiation in each GO term, to count data from the top 10,000 SNPs. Again, we used Fisher’s exact tests with Bonferroni P-value correction. From this analysis, we selected a list of the top 20 functional categories potentially enriched for each trait. We assigned P values to our comparisons of count data but used these only as a guide to select the top categories, rather than as a definitive test of enrichment. All annotation was conducted using the R library biomaRt from Bioconductor.

Code availability. Full computer code for the analysis and the values used to produce the figures are available from the lead author upon request.