Geographic Variation and Bias in the Polygenic Scores of Complex Diseases and Traits in Finland

Sini Kerminen,[1] Alicia R. Martin,[2,3,4] Jukka Koskela,[1] Sanni E. Ruotsalainen,[1] Aki S. Havulinna,[1,5] Ida Surakka,[1,6] Aarno Palotie,[1,2,3,7,8] Markus Perola,[1,5] Veikko Salomaa,[5] Mark J. Daly,[1,2,3,4] Samuli Ripatti,[1,9] and Matti Pirinen[1,9,10,*]

Polygenic scores (PSs) are becoming a useful tool to identify individuals with high genetic risk for complex diseases, and several projects are currently testing their utility for translational applications. It is also tempting to use PSs to assess whether genetic variation can explain a part of the geographic distribution of a phenotype. However, it is not well known how the population genetic properties of the training and target samples affect the geographic distribution of PSs. Here, we evaluate geographic differences, and related biases, of PSs in Finland in a geographically well-defined sample of 2,376 individuals from the National FINRISK study. First, we detect geographic differences in PSs for coronary artery disease (CAD), rheumatoid arthritis, schizophrenia, waist-hip ratio (WHR), body-mass index (BMI), and height, but not for Crohn disease or ulcerative colitis. Second, we use height as a model trait to thoroughly assess the possible population genetic biases in PSs and apply similar approaches to the other phenotypes. Most importantly, we detect suspiciously large accumulations of geographic differences for CAD, WHR, BMI, and height, suggesting bias arising from the population's genetic structure rather than from a direct genotype-phenotype association. This work demonstrates how sensitive the geographic patterns of current PSs are for small biases even within relatively homogeneous populations and provides simple tools to identify such biases. A thorough understanding of the effects of population genetic structure on PSs is essential for translational applications of PSs.

Introduction

Understanding the causes behind geographic health differences can help to optimally apply limited healthcare resources and improve public health. Geographic health differences can be partially explained by lifestyle and environmental factors, but also by genetic differences that affect health both through population-specific genetic diseases, e.g., the Finnish Disease Heritage (see Web Resources), and through variation in the polygenic components of many complex diseases.[1–3] In particular, recent discoveries from genome-wide association studies (GWASs)[4] have enabled improved polygenic prediction of complex diseases and traits and raised expectations for their future translation to clinical use.[5–8] It is an open question to what extent the geographic distribution of phenotypes could be explained by their polygenic predictions.

A standard way to estimate a polygenic score (PS) of an individual is to select a set of independent variants identified by a GWAS, to weight the number of copies of each variant by its estimate of effect size from the GWAS, and to sum these quantities over the variants. PSs have turned out to be a useful tool for identifying individuals at high risk for many diseases, such as breast cancer,[5] prostate cancer,[9] and Alzheimer disease.[8] As an example, a PS for coronary artery disease (CAD) can characterize individuals

who have a risk that is equivalent to that of carrying a monogenic variant of familial hypercholesterolemia.[7] At the same time, two recent studies have raised concerns about comparing PSs between populations with varying demographic histories.[10,11] Both studies showed that when a PS was built on a GWAS conducted in European populations and then applied to populations from Africa or East Asia, the differences in the PS were inconsistent with the actual phenotypic differences between the populations. Exact reasons for this inconsistency are unclear, but it has been speculated that a complex interplay of population genetic differences, including varying linkage-disequilibrium patterns and allele-frequency differences, between the target sample and the GWAS data can limit generalizability across populations.[10,11] Can similar problems appear also within a much more genetically and environmentally homogeneous setting than between populations from different continents? This is a crucial question for the public healthcare systems in countries that have the growing potential to implement PSs as part of their population-wide practice.

In this work, we evaluate the geographic distribution of the PSs of several complex diseases and traits in Finland and demonstrate how the effect of genetic population structure needs to be assessed before PSs can become a robust tool for population-wide use. The data resources available in

[1]Institute for Molecular Medicine Finland, Helsinki Institute of Life Sciences, University of Helsinki, Helsinki 00014, Finland; [2]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; [3]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [4]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, 02142, USA; [5]National Institute of Health and Welfare, Helsinki 00271, Finland; [6]Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA; [7]Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA 02114, USA; [8]Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA; [9]Department of Public Health, University of Helsinki, Helsinki 00014, Finland; [10]Helsinki Institute for Information Technology and Department of Mathematics and Statistics, University of Helsinki, Helsinki 00014, Finland
*Correspondence: matti.pirinen@helsinki.fi
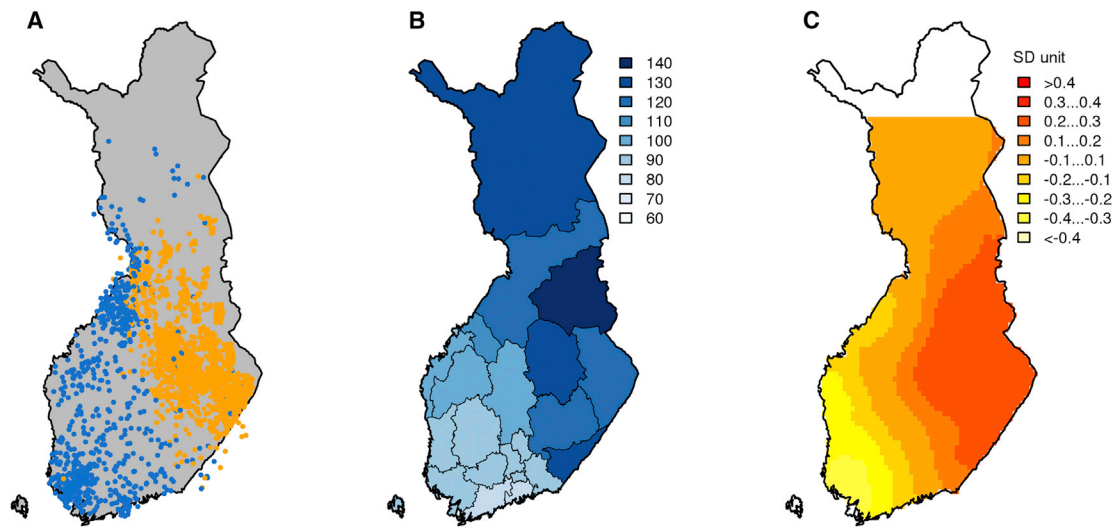https://doi.org/10.1016/j.ajhg.2019.05.001.

Figure 1. A Comparison of Genetic Population Structure, Incidence Rates, and Distribution of the Polygenic Score of Coronary Artery Disease in Finland

(A–C) Main genetic population structure (A), the incidence rate for age-adjusted coronary artery disease (CAD) in 2013–2015 (Sepelvaltimotauti-indeksi, see Web Resources) (B), and the distribution of the polygenic score (PS) for CAD (C) in Finland. The population structure was estimated by clustering 2,376 samples into two groups.[13] The incidence rate is scaled to have a mean = 100. The PS distribution is shown in units of standard deviation.

Finland provide several favorable characteristics for this study. First, on a world-wide scale, Finland has a demographically and socially homogeneous population and a top-level public healthcare system,[12] and these together reduce many possible environmental effects contributing to geographic variation in health. Second, some notable geographic differences in phenotypes and general health still occur in Finland. A good example is the CAD incidence rate that is 1.6 times higher in eastern Finland than in western Finland (Sepelvaltimotauti-indeksi, see Web Resources) (Figure 1). In fact, even larger differences in CAD incidence were observed in the 1970s, and despite the extensive and successful public health campaign to reduce these rates through the Northern Karelia project,[14] differences between east and west still remain today. Third, the genetic structure in Finland is well-characterized[13,15–19] (Figure 1), a factor that enables a detailed comparison between the geographic distribution of PSs and the overall genetic population structure within the country.

In our analyses, we observe clear geographic structure in PS distributions for most phenotypes considered. Furthermore, the spatial pattern is similar across the phenotypes and resembles the population genetic east-west division of Finland (see a comparison for CAD in Figure 1). Although a population genetic difference can well result in such patterns, a major goal of this work was to thoroughly assess whether these geographic patterns could alternatively result from some bias that emerges when the GWAS estimates of tens of thousands of variants are accumulated into PSs. We do this by generating many versions of PSs with different variant-inclusion criteria and by monitoring how the geographic structure accumulates across these PSs.

To demonstrate our approach, we consider the adult height (HG) as a model trait. In addition to HG, we apply our approach to two additional quantitative traits, body-mass index (BMI) and waist-hip ratio (WHR), in five diseases: coronary artery disease (CAD), rheumatoid arthritis (RA), schizophrenia (SCZ), Crohn disease (CD), and ulcerative colitis (UC). The results suggest that polygenic components of CAD, RA, SCZ, WHR, BMI, and HG show differences along the east-west direction, whereas only HG and WHR also show differences in the north-south direction. PSs for CD and UC do not show significant regional differences in either direction. Last, we discuss the credibility of the observed geographic differences. In particular, we report possible population-stratification-related biases in PSs for CAD, WHR, BMI, and HG. Our results raise concerns about how to reliably interpret geographic variation in PSs even within relatively homogeneous populations.

Material and Methods

Geographically Defined Target Data

We used data from the National FINRISK Study, which is a survey of the Finnish adult population (aged from 25 to 74) to estimate risk and protective factors of chronic diseases.[20] The FINRISK Study has collected several thousand samples every five years since 1972. We used data from the FINRISK Study survey of 1997 from 2,376 individuals in a geographically defined sample that was previously described in Kerminen et al.[13] The two parents of each individual in this sample were both born within 80 km of each other. For the genetic analyses, we used genotypes from Illumina HumanCoreExome-12 BeadChip (see details in Kerminen et al.[13]) and imputed genotypes as described by Ripatti et al.[21]

Table 1. Summary of the Background GWAS and Our Polygenic Scores

Trait	Study	Method	GWAS Ancestry	GWAS N	Finnish Samples	Filtering in PS	SNPs in PS
CAD	CARDIoGRAMplusC4D (Nikpay et al., 2015[23])	Logistic	European + South Asian + East Asian	60,801/123,504	5,825/5,639	P value, MAF, MHC, INFO, #Cohorts	19,597
RA	Okada et al., 2014[24]	Logistic	European	18,136/49,724	–	P value, MAF, MHC	32,736
CD	IIBDGC (Liu et al., 2015[25])	Logistic	European	5,956/14,927	–	P value, MAF, MHC, INFO, #Cohorts	21,771
UC	IIBDGC (Liu et al., 2015[25])	Logistic	European	6,968/20,464	–	P value, MAF, MHC, INFO, #Cohorts	23,513
SCZ[a]	PGC (Ripke et al., 2014[26])	Logistic	European + East Asian	36,989/113,075	–[a]	P value, MAF, MHC, INFO, #Cohorts	30,311
WHR	GIANT (Locke et al., 2015[27])	Linear	European	224,459	~16,000	P value, MAF, MHC, #Samples	13,727
	FINRISK[20]	Linear	Finnish	24,919	24,919	P value, MAF, MHC	43,252
BMI	GIANT (Shungin et al., 2015[28])	Linear	European	322,154	~23,000	P value, MAF, MHC, #Samples	12,742
	UKBB (Neale lab[29])	Linear	White British	337,199	–	P value, MAF, MHC	75,979
	FINRISK[20]	Linear	Finnish	24,919	24,919	P value, MAF, MHC	44,920
HG	GIANT (Wood et al., 2014[30])	Linear	European	253,288	~23,000	P value, MAF, MHC, #Samples	27,066
	UKBB (Neale lab[29])	Linear	White British	337,199	–	P value, MAF, MHC	113,079
	FINRISK[20]	Linear	Finnish	24,919	24,919	P value, MAF, MHC	50,536

For diseases, GWAS N = affected individuals/controls. The filtering column describes the thresholds used as follows: p value = p value < 0.05; MAF = Minor allele frequency > 0.01; MHC = major histocompatibility complex removed; INFO = imputation quality > 0.9; #Cohorts = exists in over 90% of GWAS cohorts; #Samples = exists in over 90% of GWAS samples. Other abbreviations are as follows: CAD = coronary artery disease; RA = rheumatoid arthritis; CD = Crohn disease; UC = ulcerative colitis; SCZ = schizophrenia; WHR = waist-hip ratio; BMI = body-mass index; and HG = height.
[a]Our SCZ-PS excludes Finnish samples (546/2,011) from Ripke et al. In Figures S6 and S7 we show SCZ-PS, including these Finnish samples.

Variant Filtering for Polygenic Scores

We derived PSs in our target data for each disease and trait on the basis of large international GWAS meta-analyses whose summary statistics were publicly available. We derived all PSs by excluding variants whose minor allele frequency (MAF) was below 1% in a meta-analysis, whose meta-analysis p value was above 0.05, or that resided in the major histocompatibility complex (chr 6: 25–34 Mb).[22] In addition, where applicable, we filtered out variants whose imputation quality was below 90% or variants that had been present in less than 90% of the cohorts or samples of the meta-analysis. We also excluded all multi-allelic variants. Table 1 summarizes GWAS characteristics and variant filtering for all PSs. Finally, the PSs were built by selecting independent variants with PLINK 1.9[31] (see Web Resources) via the clump command with a 500 kb window radius and a 0.1 threshold for r^2.

Additional GWASs for FINRISK, UK Biobank, and GIANT

FINRISK

We used Hail (see Web Resources) to run standard linear regression for HG, BMI, and WHR (adjusted for BMI) in 24,919 individuals across the National FINRISK Study collections from 1992–2012. These data excluded all 2,376 target individuals. We used sex, age, FINRISK project year, genotyping chip, and the first ten principal components of population structure as covariates in the analysis. In addition, we ran a linear mixed model for HG with BOLT-LMM v.2.3[32] with the same samples and covariates.

UK Biobank

For the UK Biobank (UKBB), we performed a linear mixed model GWAS for HG with BOLT-LMM v2.3.[32] For this analysis, we mimicked the linear regression analysis (round 1) of the Neale lab[29] and used UKBB v2 genotypes on 343,728 samples with white British ancestry. We used age, sex, and the first 20 principal components as covariates, and we used directly genotyped variants with a MAF above 1% and missingness below 10% for generating the variance component. GWAS statistics were calculated for imputed data with a MAF above 0.1% and an imputation quality above 0.7.

GIANT

We made two additional versions of the GIANT consortium meta-analysis with METAL,[33] as in Wood et al.,[30] except that in the first version we excluded the cohorts that included samples from the National FINRISK Study (FUSION, MIGEN, and COROGENE) and in the second version we excluded all cohorts that included any Finnish samples (DGI, FTC, FUSION, GENMETS, MIGEN, NFBC1966, COROGENE, FINGESTURE, HBCS, and YFS).

Polygenic Scores

We calculated PSs for the target set of 2,376 FINRISK individuals by using the additive model of:

$$PS_i = \sum_{j=1}^{M} x_{ij}\hat{\beta}_j,$$

Table 2. Results from the Linear Model for Correlated Data where Polygenic Score is Explained by Latitude or Longitude

| Trait | SNPs | Latitude | | Longitude | | WF-EF Difference (95% CI) |
		Estimate	P value	Estimate	P value	
CAD	19,597	-6.3×10^{-4}	0.97	0.05	1.6×10^{-4}*	−0.63 (−0.71, −0.55)
RA	32,736	0.03	0.12	0.06	5.5×10^{-5}*	−0.63 (−0.71, −0.55)
CD	21,771	0.03	0.18	−0.002	0.87	0.10 (0.01, 0.19)
UC	23,513	0.03	0.23	0.02	0.22	−0.26 (−0.35, −0.18)
SCZ	30,311	0.04	8.7×10^{-2}	0.04	4.0×10^{-3}*	−0.35 (−0.43, −0.26)
BMI	12,742	0.03	9.4×10^{-2}	0.04	1.8×10^{-3}*	−0.53 (−0.61, −0.44)
WHR	13,727	0.10	1.0×10^{-9}*	0.08	4.7×10^{-12}*	−1.16 (−1.23, −1.09)
HG	27,066	−0.18	1.1×10^{-40}*	−0.15	2.1×10^{-60}*	1.51 (1.45, 1.58)

SNPs = number of variants in PS. The difference in PS between Western Finland (WF) and Eastern Finland (EF) subpopulations is given in the standard deviation unit of PS. * marks a p value < 0.05. Abbreviations are as follows: CI = confidence interval, CAD = coronary artery disease; RA = rheumatoid arthritis; CD = Crohn disease; UC = ulcerative colitis; SCZ = schizophrenia; WHR = waist-hip ratio; BMI = body-mass index; and HG = height.

where PS_i is a polygenic score for individual i, M is the number of SNPs in the score (after variant filtering), x_{ij} is the individual's (imputed) genotype dosage for SNP j, and $\widehat{\beta}_j$ is the effect size estimate of SNP j from the GWAS.

Genetic Risk Maps

To visualize the geographic distribution of PSs, we used the geographic locations of our geographically well-defined sample of 2,376 individuals and their PSs. We estimated an individual's geographic location as the mean of his or her parents' birthplaces. We then created risk maps in R by using a geographical centroid approach; this approach lays a grid on the map of Finland, and for each grid point p, it calculates the average of individuals' PSs inversely weighted by their squared distance to the grid point as

$$PS_p = \frac{1}{r_{Tot}} \sum_{i=1}^{N} \frac{PS_i}{r_{ip}^2},$$

where r_{ip} is the distance between individual i and grid point p and $r_{Tot} = \sum_i 1/r_{ip}^2$ is the sum of the weights. We used a grid with a square size of 10 km and limited the minimum value for r_{ip} to be 50 km to avoid high variance in weights. In addition, to control for uncertainty in the areas that have a low sample size, we added to the calculation of PS_p one pseudo-individual whose PS is the population average PS and whose distance to the point p is the minimum of the observed distances r_{ip}. This modification draws the PS values of grid points toward the population average, especially in sparse areas where there are few individuals at the minimum range from the grid point. Last, the risk maps were scaled by the population average and standard deviation with a subset of 1,042 geographically evenly distributed individuals as described in Kerminen et al.[13] The border line for the map of Finland was obtained from GADM (see Web Resources).

A Linear Model for Correlated Data for Assessing Spatial PS Differences

To quantify whether the PS has geographic differences, we performed a regression analysis with a linear model for correlated data wherein we explained PS with latitudinal or longitudinal coordinates and accounted for genetic relatedness as

$$PS_i = \mu + x_i\alpha + \varepsilon \quad \varepsilon \sim N(0, \sigma_\varepsilon^2 R),$$

where x_i is the coordinate of individual i, μ is the intercept, and α is the effect of latitude or longitude on PS reported in Tables 2, S4, S6 and S7. For the structure of the error terms, we used the genetic relationship matrix R that was estimated with PLINK 1.9[31,34] (command --make-rel) with 61,598 independent variants from the Illumina HumanCoreExome chip described in Kerminen et al.[13] Regression results from the standard linear model without accounting for genetic relatedness are shown in Table S1.

Polygenic and Phenotypic Differences between Subpopulations

The two main subpopulations in Finland are located in western Finland (WF) and eastern Finland (EF); they were previously described in Kerminen et al.[13] and are shown in Figure 1A. Here, we reproduced this analysis by using CHROMOPAINTER and FineSTRUCTURE[35] with our current sample of 2,376 individuals to estimate both phenotypic and polygenic score differences between these two populations. The analysis divided our target sample into 1,604 EF and 772 WF individuals, and we used this division for estimating the differences between subpopulations.

PS Differences in Standard Deviation Units

We calculated the PS differences between the subpopulations by first scaling the PSs of the target sample with the subset of 1,042 geographically evenly distributed samples. We then used scaled PSs to calculate the difference between WF and EF. This strategy ensured a robust comparison between PSs on the basis of a fixed reference set. The 95% confidence intervals for the difference between two groups were given by Welch's t test in R 3.4.1.[36]

Phenotypic Differences Predicted by PS

For HG, BMI, and WHR, we also estimated the phenotypic difference predicted by PS between our subpopulations. First, we fitted the linear model wherein we explained the phenotype with the general covariates of sex, age (measured in 1997 and also represented by birth year), and age^2 (WHR was additionally adjusted for BMI) in our target sample. Then we fitted another

linear model wherein we explained the residuals with PS. On the basis of the effect estimates of the second model, we were able to estimate the predicted phenotypic effect by multiplying the PS effect estimate with the PS difference between the populations. We estimated the respective 95% credible intervals by using a simulation approach wherein we generated 100,000 sample pairs of effect estimates for PS difference, d, and PS effect on phenotype, β, from their sampling distributions, and we used the empirical distribution of d*β to determine the 95% credible interval. The sampling distribution of d was modeled as a normal distribution with a mean set to the observed PS difference and the standard deviation calculated from the 95% confidence interval from the Welch's t test as $(\overline{x} - CI_{low})/1.96$. The sampling distribution of β was modeled as a normal distribution with a mean set to the observed effect estimate and the standard deviation set to the corresponding standard error from the linear model.

Observed Phenotypic Differences for HG, BMI, and WHR

We estimated the observed phenotypic differences in HG, BMI, and WHR between WF and EF by adjusting the corresponding trait for sex, age, and age^2 (WHR was additionally adjusted for BMI) via linear regression, and then we calculated the difference of the subpopulation means on the basis of the residuals from this regression. The residuals were maintained in the units of the original phenotypes.

P Value Thresholding in PSs

We studied the effect of p value threshold for our PSs by applying seven different thresholds (p value $< 1 \times 10^{-2}$, 1×10^{-3}, 1×10^{-4}, 1×10^{-5}, 1×10^{-6}, 1×10^{-7}, and 1×10^{-8}) to the variants of initial PSs (that used a threshold of 0.05) and calculated the additional PSs as described above.

Detecting Accumulation of Biases with Weakly Associated Variants ("Random PSs")

To detect the accumulation of biases, we used an approach where we first filtered the GWAS summary statistics similarly to the original scores (as explained above), except we considered only variants with the GWAS p value > 0.5. This left us with, at most, very weakly associated variants. Among these p > 0.5 variants, we performed linkage disequilibrium (LD) clumping with the same parameters as above, except we set the p value cut-off to 1 in order to not further exclude any variant on the basis of its p value, and we permuted the p values among the p > 0.5 variants to ensure that the resulting scores are random with respect to their p value. LD clumping resulted in different numbers of variants for different traits, and from those we randomly sampled increasing numbers of variants (5,000, 10,000, 20,000, 40,000, 60,000, and 80,000). For BMI and WHR, the remaining number of variants after LD clumping was <80,000, and hence we were not able to compute a PS for 80,000 for these two traits. Finally, we calculated the PS for each individual and evaluated the difference between subpopulations in these "random PSs." To assess uncertainty, we repeated the random PS generation ten times, and we report the mean and the range of the subpopulation difference over these ten random PSs.

To understand the expected behavior of PSs with truly zero effect sizes and to compare with our observed random PSs, we generated 1,000 simulated PSs for each observed random PS. These PSs were simulated from the variants from the random PSs but sampled their effect estimates independently from a normal distribution with a mean of zero and a standard deviation that corre-

sponded to the standard error of the variant in GWAS. In Figure 5, we see that the 95% highest probability interval of the population difference is approximately constant across the different number of variants in the PSs and across the different GWASs. Supplemental Text S1 describes the theoretical basis for this property.

Our simulated 95% intervals assume completely independent variants, whereas our PS pipeline used a more liberal LD threshold of $r^2 < 0.1$. Therefore, we also compared the effects of residual LD to our random scores by performing LD clumping with r^2 thresholds of 0.01 and 0.001 for CAD and HG with GIANT and UKBB data. Figure S1 shows that, for GIANT-PS, the residual LD does not have an effect on the accumulation of population difference, and a similar tendency is suggested for CAD-PS even though the data are limited. For UKBB-PS, there is no accumulation of difference for any r^2 threshold.

The imputation-quality filter did not noticeably affect the results from either the actual PS or the random PS (Figure S2).

Results

Polygenic Scores Show Geographic Differences in Finland

We estimated PSs across Finland by using a geographically well-defined sample of 2,376 individuals from the National FINRISK Study 1997 survey.[20] The parents of each of these 2,376 individuals were born within 80 km of each other, and the means of the parents' coordinates were used as the individuals' locations. We derived PSs for the individuals with summary statistics from publicly available GWAS meta-analyses by applying LD pruning ($r^2 < 0.1$), MAF filtering (MAF > 0.01), and p value thresholding ($p < 0.05$) (see Material and Methods). To visualize the results on the map of Finland (Figure 2), we estimated the score at each map point by averaging individuals' PSs inversely weighted by the individuals' squared distance from the point (see Material and Methods).

We applied our approach to five diseases, CAD, RA, CD, UC, and SCZ, as well as to three quantitative traits, BMI, WHR adjusted for BMI, and HG. We observe that the PS patterns for CAD, RA, SCZ, BMI, HG, and WHR closely resemble the main population structure in Finland (Figure 1A). CD and UC do not show clear geographic differences between any parts of the country.

To evaluate statistically whether the PSs show geographic differences, we quantified the patterns by using a linear model for correlated data, wherein we explained individuals' PSs with either longitude or latitude and accounted for genetic relatedness of the samples (see Material and Methods). The strongest differences were observed for HG ($p = 2.1 \times 10^{-60}$) and WHR ($p = 4.7 \times 10^{-12}$) based on longitude, and lower but non-zero differences were observed for CAD, RA, SCZ, and BMI (all with $p < 0.05$) (Table 2, see Table S1 for results based on the standard linear model without accounting for genetic relatedness). HG and WHR showed differences also based on latitude, whereas CD and UC did not show differences based on either longitude or latitude. Table 2 also shows that the difference in PS between WF and EF subpopulations is
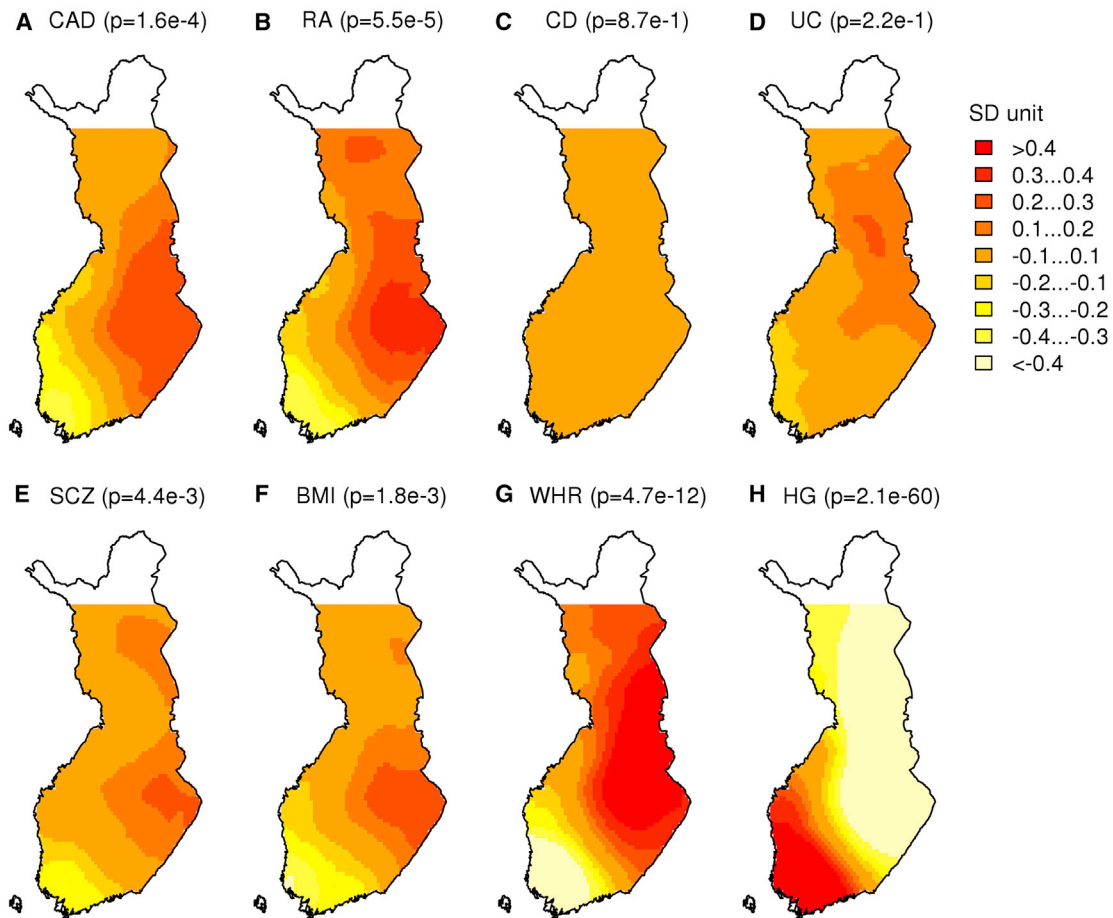
Figure 2. Distribution of Polygenic Scores in Finland
(A–H) Distribution of polygenic scores for (A) coronary artery disease, (B) rheumatoid arthritis, (C) Crohn disease, (D) ulcerative colitis, (E) schizophrenia, (F) body-mass index, (G) waist-hip ratio adjusted for body-mass index, and (H) height. P values correspond to the association with longitude presented in Table 2.

the largest in HG (1.51 SDs) and in WHR (−1.16 SDs). In general, we observed stronger PS differences between east and west than between north and south, and this observation is in line with the main population structure in Finland.

Recently, it has been reported that PS differences between populations are prone to technical and confounding biases arising especially from population genetic differences (i.e., genetic divergence) or relatedness structure between the GWAS discovery and the target data.[10,11,37–39] To assess whether some of the results in Figure 2 and Table 2 might be affected by these problems, we next concentrate on evaluating our PSs in several ways. We used HG as a model trait for developing the methodology.

Height PSs in Three Independent Cohorts
Adult height (HG) is a highly heritable and polygenic trait[30,40,41] and shows clear phenotypic differences in Finland; western Finns are on average 1.6 cm taller than eastern Finns (see Material and Methods and Figure 3A). Furthermore, HG is a quantitative trait that makes it possible to compare geographic differences between the

observed phenotype and the predictions based on PS. For such comparisons, we regressed out effects of sex, age, and age^2 from HG by using residuals from a standard linear model.

We calculated HG-PS by using summary statistics from three independent GWASs, including results from the GIANT consortium (a meta-analysis from a heterogeneous set of European samples),[30] the UK Biobank (a single cohort of uniformly genotyped and phenotyped white-British samples), and the National FINRISK study (Finnish samples genotyped with two different chips) using our standard pipeline (see Material and Methods). Table 3 summarizes the performance of these three scores.

The GIANT consortium height GWAS is a meta-analysis of 250,000 samples from multiple European populations, and it includes about 23,000 Finnish samples.[30] The GIANT-PS included 27,000 variants, explained 14% of the variance of height, and showed dramatic geographic differences in Finland (Figure 3B). The GIANT-PS was 1.5 SDs larger in WF than in EF, and we estimated, by regressing height on this PS in the target sample of 2,376 Finnish individuals, that this difference would correspond to a
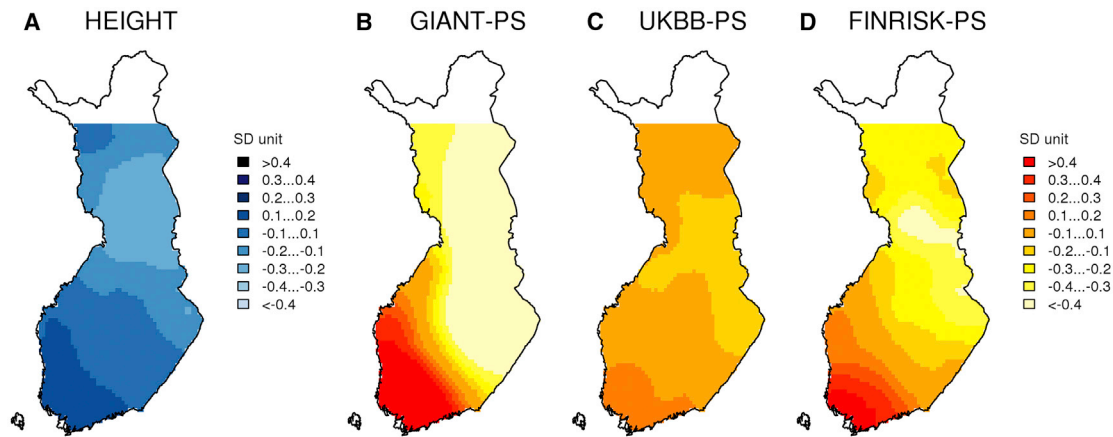
Figure 3. Distributions of Adult Height and Three Polygenic Scores for Height
(A–D) A distribution of sex-, age-, and age²-adjusted adult height (A) and polygenic score (PS) distributions of GIANT-PS (B), UKBB-PS (C), and FINRISK-PS (D) for height in Finland. The values are presented in standard deviation units.

predicted height difference of 3.5 cm between WF and EF (see Material and Methods). This difference is over twice the observed phenotypic difference between the subpopulations. Note that even if we assumed that all variation in height was genetic, we would expect our GIANT-PS (that has $R^2 < 15\%$) to explain only a part of the actual 1.6 cm WF-EF height difference. This raises concerns that in our target sample, GIANT-PS produces geographically biased results that cannot be interpreted directly on the phenotypic scale. The predicted WF-EF difference was even larger for GIANT-PS if the HG-on-PS regression was done within the WF subpopulation (4.7 cm) or within the EF subpopulation (6.4 cm) alone (Table S2), indicating challenges of interpretability for absolute differences among subpopulations.

Second, we built a PS based on over 330,000 UK Biobank British ancestry samples analyzed by the team led by Benjamin Neale.[29] Using the same pipeline as with GIANT-PS, this UKBB-PS contained considerably more variants and gave qualitatively similar geographic results to GIANT-PS, but quantitatively it showed much smaller WF-EF differences (Figure 3C). UKBB-PS explained 22% of the variation of height in the target sample and corresponded to a 0.6 cm predicted WF-EF difference in height.

Third, we built a PS based on the Finnish-population-specific summary statistics from the National FINRISK Study.[20] This FINRISK GWAS included nearly 25,000 samples and excluded all our 2,376 target individuals. This FINRISK-PS (50,000 SNPs) explained 15% of the variance of height and showed significant WF-EF differences that corresponded to a 1.4 cm difference in predicted height (Figure 3D). For FINRISK-PS and UKBB-PS, the predictions were robust to whether the regression was done in the whole target sample or in its WF or EF subset alone (Table S2).

These results show a consistent direction in predicted height differences between western and eastern Finland on the basis of three independent GWASs that have

different relationships to the target sample. The predicted direction is also consistent with the observed phenotypic difference. However, the results show considerable, and concerning, variation in the predicted geographic difference of the genetic component of height.

Evaluating Possible Biases in Polygenic Score for Height
A PS from GIANT Accumulates Geographic Differences
An accumulation of small biases might be a substantial risk in the PSs of thousands of variants. These small biases can arise, for example, from unadjusted population structure in the underlying GWAS or from overlapping samples between GWAS and target data.[42,43] To understand whether the differences in our PS and the unrealistic predictions of geographic height differences might be due to a bias accumulation, we generated additional PSs by varying the inclusion criteria of variants.

First, we used variants from the initial PS but applied different p value thresholds. Even though these scores included different numbers of variants, the variance explained did not vary strongly across the thresholds for GIANT-PS or UKBB-PS (Figure 4A), a finding that replicates the behavior of PSs reported earlier by Wood et al.[30] for other target populations. For FINRISK, the variance explained increased considerably with a more liberal p value threshold because of the smaller sample size of the study; specifically, the FINRISK-PS included only a handful of variants for the smallest p value thresholds, and thus had only a little predictive power there (Figure 4A). Conversely, the predicted east-west height differences in the GIANT-PS decreased considerably as more stringent p value cutoffs were used, whereas the variance in height explained by the PS increased simultaneously (Figure 4B). The decrease was much subtler for the other two PSs. To confirm the effect of the number of variants on the predicted height differences, we randomly sampled 1,000 variants from each of the different p value thresholds in GIANT-PS and calculated the corresponding scores. These PSs showed

Table 3. Summary of the Results in the HG-PS Comparison

Source GWAS	Ancestry	GWAS N	Finnish Samples	Variants in PS	Adjusted R^2	Predicted WF-EF HG Difference (cm; 95% CI)	Observed WF-EF HG-PS Difference (SD unit; 95% CI)
GIANT	European	253,288	~23,000	27,066	14%	3.52 (3.14, 3.90)	1.51 (1.45, 1.5)
GIANT NOFINNS	European	230,794	0	25,660	17%	1.78 (1.53, 2.05)	0.70 (0.62, 0.79)
UK Biobank	British	337,199	0	113,079	22%	0.64 (0.39, 0.89)	0.23 (0.14, 0.32)
FINRISK	Finnish	24,919	24,919	50,536	15%	1.35 (1.14, 1.58)	0.59 (0.51, 0.67)

Adjusted R^2 is the variance explained by the PS in the target set.

similar levels of predicted WF-EF height difference (about 1 cm) independent of the p value threshold, suggesting that the number of variants is a more important factor behind the geographic structure of GIANT-PS than the actual phenotypic variance explained by the variants (Figure S3).

Concerned about the accumulation of WF-EF differences in GIANT-PS, we tested whether similar accumulation occurred even over random, non-associated variants. We randomly sampled different numbers of independent variants whose p values were over 0.5 in GWAS (suggesting a negligible association to height) and calculated PSs for them (we call these "random PSs"). This test showed considerable geographic differences in random PSs based on GIANT GWAS, and these differences increased with the number of variants (Figures 4C and 4D). Similar but weaker behavior was detected for FINRISK variants but not for UKBB variants.

Correlation of PS with Principal Component 1
One potential explanation for the observed behavior of GIANT-PS is that the effect-size estimates have a consistent directional bias that is aligned with the main population structure in Finland. Indeed, GIANT-PS is highly correlated with the leading principal component (PC1) of population structure in our target sample (r = 0.80), and when we removed the linear effect of PC1 from the GIANT-PS (see Material and Methods), the residuals explained more variance (19%) of height than the original GIANT-PS (14%). This suggests that in the effect-size estimates of GIANT-PS, a part of the true height association is masked by a strong component aligned with PC1 in Finland. For neither FINRISK-PS nor UKBB-PS did the removal of the linear effect of PC1 improve the variance explained in our target sample (Table S3). None of the three PSs showed WF-EF differences after the PC1 was regressed out (Table S3). By using the variants from GIANT-PS and the effect estimates from UKBB, we observed that the strong geographic differences in GIANT-PS are likely driven also by the choice of the GIANT variants and not only by a bias in the GIANT effect estimates (Supplemental Text S2).

Together, these analyses suggest that the geographic distribution of PSs based on the GIANT summary statistics consistently exaggerates height differences between the main Finnish subpopulations, whereas much less confounding from population stratification is seen in FINRISK-PS, and almost none is observed in UKBB-PS. A few possible reasons for this bias accumulation could be the inadequate adjustment for population structure in GWASs[37,39] or partially overlapping or related samples between GWAS samples and test data.

Effect of Sample Overlap and Population-Specific Samples
Our target data originate from the National FINRISK Study that is not reported among the GIANT cohorts (neither among cohorts in Lango et al.,[44] nor among additional cohorts in Wood et al.[30]). However, a closer look into the cohort descriptions suggested that the COROGENE, FUSION, and MIGEN cohorts might include some FINRISK samples. This shows that it is not always straightforward to keep track of where publicly accessible samples have been used previously, and this observation would be a crucial piece of information for appropriately validating PSs. Although computational methods exist to detect sample overlaps between GWAS summary statistics,[45] their behavior in large datasets is not yet completely understood.[46]

To test whether overlapping individuals affect the results of GIANT-PS, we built a new PS without any FINRISK samples (called GIANT-NOFR-PS). This GIANT-NOFR-PS explained 16% of the height variance and predicted a 3.4 cm (95% CI: 3.0–3.7) difference between WF and EF, and this difference is very similar to the result of the original GIANT-PS (14%, 3.5 cm [95% CI: 3.2–3.9]). Thus, a possible sample overlap is not causing the exaggerated WF and EF differences in GIANT-PS.

Next, we excluded the cohorts that included any Finnish samples from the meta-analysis. This GIANT-NOFINNS-PS explained 17% of the height variance and predicted a 1.7 cm (95% CI: 1.5–2.1) difference between WF and EF (Figure S4). This is a significantly smaller difference than that found by the original GIANT-PS, and a similar drop is also seen in the bias accumulation of GIANT-NOFINNS-PS (Figures 4 and S5). A similar but weaker effect was detected for SCZ when comparing PSs based on meta-analyses with and without Finnish samples (see Figures S6 and S7 and Table S4). These results suggest that although population-specific PSs have the potential to increase prediction accuracy, they might also introduce considerable bias if the population structure has not been adjusted properly.
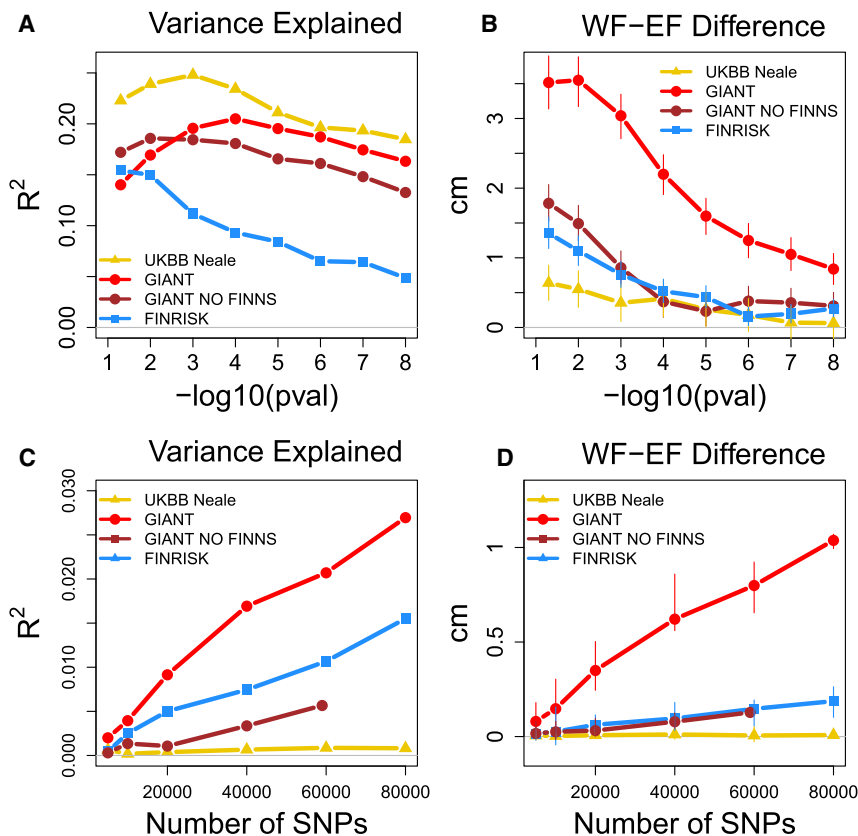
Figure 4. A Comparison of PSs Constructed from Height-Associated versus Random SNPs Demonstrates Differences in Stratification Effects by GWAS Summary Statistics

Top row: (A and B) Height variance explained by PS (A) and east-west difference in height predicted by PS (B) as a function of p value threshold in GWAS data.

Bottom row: (C and D) Height variance explained by PS (C) and east-west difference in height predicted by PS (D) as a function of the number of independent variants in PS when all variants have a p value > 0.5 in GWAS ("random scores"). Variance explained is given as adjusted R^2. In C and D, the values are based on ten random scores, and error bars in D show the range of those scores.

Discussion

Polygenic scores (PSs) have recently reached predictive power for some well-established monogenic risk factors for disease,[7] and several projects are currently testing their utility in health care settings. PSs could also potentially inform us about the role of genetics in the geographic variability of traits and disease. However, a major challenge is that the geographic distribution of PSs is a complex function of population genetic differences between the GWAS data and the target samples, complicating its interpretation.[10,11,47,48] Here, we studied the geographic distribution of several PSs within Finland and assessed their robustness and possible biases in several ways.

By generating PSs for eight phenotypes in Finnish samples, we observed strong similarities between the geographic distribution of several PSs and the main Finnish population structure that runs from south-west to north-east.[13] We further showed that even the least statistically significantly associated half of the effect sizes (with GWAS p value > 0.5) was carrying a consistent pattern of east-west difference for CAD (CARDIoGRAM data) and the three anthropometric traits from the GIANT consortium: HG, BMI, and WHR, findings that we interpret to indicate a likely bias. In theory, such a pattern could also result from extreme polygenicity. However, with the highly polygenic HG as our model trait, we showed that the random score from our largest HG GWAS based on the UK Biobank did not show any east-west variation within Finland. This suggests that the geographic difference accumulating in the random score from GIANT is rather due to bias than polygenicity. Furthermore, we observed for HG that the GIANT-PS was so strongly aligned with the first principal component of the genetic structure in our target data that this association masked some of the predictive power of the PS. This suggests that the effect
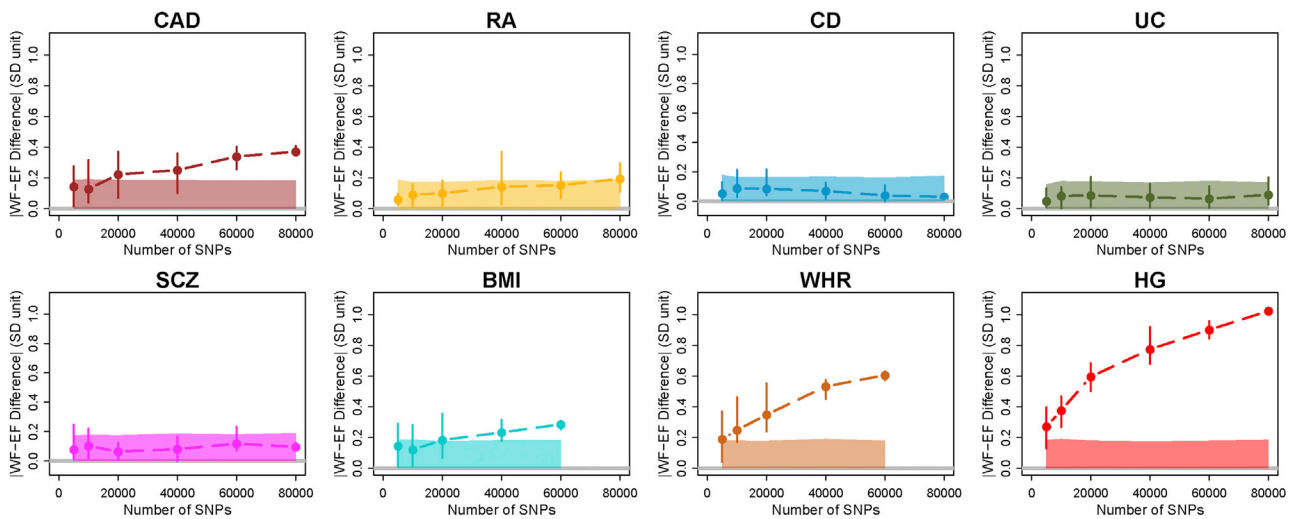
When we ran the HG-GWAS in FINRISK and UKBB data with a linear mixed model that accounts for genetic relatedness, the resulting PSs predicted smaller geographic differences between WF and EF than our original GWAS based on linear regression with ten PCs as covariates, but the differences were not significant (Figure S8 and Table S5).

Bias Accumulation in Other Complex Diseases and Traits

After assessing multiple sources of bias accumulation in HG, we applied similar strategies to the other seven phenotypes. For each phenotype, we generated random PSs with increasing numbers of variants to detect a possible accumulation of biases. Here, we present absolute difference between western and eastern Finland using standardized PSs because we did not have a way to turn these to the phenotypic scale for the disease studies. Figure 5 shows that for RA, CD, UC, and SCZ, the absolute WF-EF PS difference of the random score is close to zero, whereas for CAD, BMI, WHR, and HG we observe a possible accumulation of bias.

Similarly to HG, we were able to compare the geographic distribution of the PSs of BMI and WHR to their phenotypic counterparts. Neither BMI nor WHR shows clear geographic patterns in our data, but the PSs of BMI and WHR largely repeat the results of the PS of HG (Figures S9, S10, S11, and S12 and Tables S6 and S7). For both BMI and WHR, GIANT-PS shows much larger differences and bias accumulation in random scores between WF-EF than FINRISK-PS (BMI and WHR) or UKBB-PS (only BMI available).

Figure 5. Accumulation of Geographic Differences in Random Polygenic Scores
The absolute value of PS difference between western and eastern subpopulations with different numbers of independent variants ($r^2 < 0.1$) randomly chosen with GWAS p value > 0.5. For BMI and WHR, the data did not contain more than 60,000 independent variants. The solid region is the 95% probability interval under the theoretical null assumption of zero effect sizes and completely independent variants ($r^2 = 0$) (see Material and Methods). Points show the mean, and error bars show the range over ten random scores.

estimates from GIANT contain a bias aligned with the main population structure in Finland, and this finding is in line with two recent studies that have reported related biases in the context of polygenic selection studies.[37,39] When we removed all Finnish samples from the GIANT meta-analysis of HG, the east-west difference in PS was halved, but still remained threefold compared to that in UKBB-PS. Also, the random scores showed that although a considerable proportion of the bias in GIANT-PS was related to the Finnish GWAS samples, a considerable bias still remained after excluding the Finnish samples.

For all three quantitative traits, PSs predicted unrealistically large geographic differences compared to the actual phenotypic differences. A theoretical but unlikely possibility remains that the geographic structure of the genetic component not explained by our current PSs could be opposite to the component that is explained by our current PSs, and this could eventually balance out the unrealistically large estimates for GIANT-PS and FINRISK-PS. However, given that the estimated difference consistently increases with the inclusion of more variants in PSs, a more plausible explanation is that we simply cannot robustly interpret the geographic differences in PSs derived from existing GWASs on the phenotypic scale via a simple regression framework. Earlier, the results of phenotypically inconsistent PS differences between continental groups have been reported.[10,11,38] Here, we show that similar patterns can exist even for a relatively small geographic area and the relatively homogeneous population of Finland. We note that even if the genetic EF-WF difference in Finland might be large compared to variation within some other European countries,[19] it is tiny compared to the continental differences.[49]

Our results showed that the phenotypes that did not accumulate EF-WF differences were the two types of inflammatory bowel disease (CD and UC), SCZ, and RA. Of these, CD and UC did not show any geographic PS variation in Finland. To our knowledge, only two studies have studied the geographic variation in the prevalence of inflammatory bowel disease (IBD) in Finland. Lehtinen et al.[50] reported higher incidence rates of pediatric IBD in more sparsely populated areas, whereas Jussila et al.[51] reported increasing prevalence rates of UC in Northern Finland but no geographic structure for CD. Our polygenic risk prediction for CD is in line with the observations in Jussila et al.,[51] and even though the PS of UC did not show significant geographic differences in our statistical analysis, the genetic risk map for UC shows some increasing risk pattern in Northern Finland. SCZ showed a higher polygenic risk in EF than in WF, a finding in line with extensive geographic incidence information from several studies[52-57] that describe the highest SCZ prevalence and incidence rates in northern and eastern Finland and the lowest rates in the southwestern parts of the country. Also, RA showed higher polygenic risk in EF than in WF. Our limited information about the regional incidence of RA in Finland is from Kaipiainen-Seppänen et al.,[58] who reported the highest RA incidence rates for North Karelia (in EF) and the lowest for Ostrobothnia (on the west coast), but unfortunately the study did not include southwestern or northern Finland. Neither our SCZ nor our RA GWAS summary statistics included any Finnish samples. Together these two diseases exemplify the potential of PSs to explain geographic health differences.

To conclude, we recommend the following practices for geographic evaluation of PSs. (1) Check residual geographic stratification of PSs by generating random scores for non-associated variants and by testing whether PSs unrealistically strongly align with the leading PCs of genetic structure. (2) Use a linear/logistic mixed model instead of the standard linear/logistic regression model in GWAS. (3) Compare the genetically predicted phenotypic difference between populations to the observed phenotypic difference in order to detect unrealistic genetic predictions. With these tools, we showed that although PSs for several traits in Finland followed the geographic distribution of the phenotype (HG, CAD, SCZ, RA, CD, and UC), for CAD and HG, as well as for BMI and WHR, we observed in the geographic distribution of PSs suspicious behavior that could indicate a bias arising from population genetic structure rather than from a direct genotype-phenotype association. Our results emphasize that we have limited understanding of the interplay between our current PSs and genetic population structure even within one of the most thoroughly studied populations in human genetics. Therefore, we recommend refraining from using the current PSs to argue for a significant polygenic basis for geographic phenotype differences until we understand better the source and extent of the geographic bias in the current PSs.

Supplemental Data

Acknowledgments

Web Resources

The Finnish Disease Heritage, http://www.findis.org/
GADM, https://gadm.org/
Hail, https://github.com/hail-is/hail
PLINK 1.9., www.cog-genomics.org/plink/1.9/
Sepelvaltimotauti-indeksi, T.H.L. Ikävakioitu (2013–2015), http://www.terveytemme.fi/sairastavuusindeksi/2015/maakunnat_html/atlas.html?select=01&indicator=i0

References

1. Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D.J., et al. (2016). The genetic architecture of type 2 diabetes. Nature *536*, 41–47.

2. Hartiala, J., Schwartzman, W.S., Gabbay, J., Ghazalpour, A., Bennett, B.J., and Allayee, H. (2017). The genetic architecture of coronary artery disease: Current knowledge and future opportunities. Curr. Atheroscler. Rep. *19*, 6.

3. O'Connell, K.S., McGregor, N.W., Lochner, C., Emsley, R., and Warnich, L. (2018). The genetic architecture of schizophrenia, bipolar disorder, obsessive-compulsive disorder and autism spectrum disorder. Mol. Cell. Neurosci. *88*, 300–307.

4. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. Am. J. Hum. Genet. *101*, 5–22.

5. Mavaddat, N., Pharoah, P.D., Michailidou, K., Tyrer, J., Brook, M.N., Bolla, M.K., Wang, Q., Dennis, J., Dunning, A.M., Shah, M., et al. (2015). Prediction of breast cancer risk based on profiling with common genetic variants. J. Natl. Cancer Inst. *107*, djv036.

6. Abraham, G., Havulinna, A.S., Bhalala, O.G., Byars, S.G., De Livera, A.M., Yetukuri, L., Tikkanen, E., Perola, M., Schunkert, H., Sijbrands, E.J., et al. (2016). Genomic prediction of coronary heart disease. Eur. Heart J. *37*, 3267–3278.

7. Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., and Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat. Genet. *50*, 1219–1224.

8. Stocker, H., Möllers, T., Perna, L., and Brenner, H. (2018). The genetic risk of Alzheimer's disease beyond APOE ε4: Systematic review of Alzheimer's genetic risk scores. Transl. Psychiatry *8*, 166.

9. Schumacher, F.R., Al Olama, A.A., Berndt, S.I., Benlloch, S., Ahmed, M., Saunders, E.J., Dadaev, T., Leongamornlert, D., Anokian, E., Cieza-Borrella, C., et al.; Profile Study; Australian Prostate Cancer BioResource (APCB); IMPACT Study; Canary PASS Investigators; Breast and Prostate Cancer Cohort

Consortium (BPC3); PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium; Cancer of the Prostate in Sweden (CAPS); Prostate Cancer Genome-wide Association Study of Uncommon Susceptibility Loci (PEGASUS); Genetic Associations and Mechanisms in Oncology (GAME-ON)/Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE) Consortium (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat. Genet. *50*, 928–936.

10. Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., Daly, M.J., Bustamante, C.D., and Kenny, E.E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. Am. J. Hum. Genet. *100*, 635–649.

11. Reisberg, S., Iljasenko, T., Läll, K., Fischer, K., and Vilo, J. (2017). Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. PLoS ONE *12*, e0179238.

12. GBD 2016 Healthcare Access and Quality Collaborators (2018). Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: A systematic analysis from the Global Burden of Disease Study 2016. Lancet *391*, 2236–2271.

13. Kerminen, S., Havulinna, A.S., Hellenthal, G., Martin, A.R., Sarin, A.P., Perola, M., Palotie, A., Salomaa, V., Daly, M.J., Ripatti, S., and Pirinen, M. (2017). Fine-scale genetic structure in Finland. G3 (Bethesda) *7*, 3459–3468.

14. Puska, P., Vartiainen, E., Laatikainen, T., Jousilahti, P., and Paavola, M. (2009). The Norther Karelia Project: From North Karelia To National Action (Helsinki: National Institute for Health and Welfare (THL), in collaboration with the North Karelia Project Foundation).

15. Jakkula, E., Rehnström, K., Varilo, T., Pietiläinen, O.P., Paunio, T., Pedersen, N.L., deFaire, U., Järvelin, M.R., Saharinen, J., Freimer, N., et al. (2008). The genome-wide patterns of variation expose significant substructure in a founder population. Am. J. Hum. Genet. *83*, 787–794.

16. Lappalainen, T., Koivumäki, S., Salmela, E., Huoponen, K., Sistonen, P., Savontaus, M.L., and Lahermo, P. (2006). Regional differences among the Finns: A Y-chromosomal perspective. Gene *376*, 207–215.

17. Martin, A.R., Karczewski, K.J., Kerminen, S., Kurki, M.I., Sarin, A.P., Artomov, M., Eriksson, J.G., Esko, T., Genovese, G., Havulinna, A.S., et al. (2018). Haplotype sharing provides insights into fine-scale population history and disease in Finland. Am. J. Hum. Genet. *102*, 760–775.

18. Neuvonen, A.M., Putkonen, M., Översti, S., Sundell, T., Onkamo, P., Sajantila, A., and Palo, J.U. (2015). Vestiges of an ancient border in the contemporary genetic diversity of North-Eastern Europe. PLoS ONE *10*, e0130331.

19. Salmela, E., Lappalainen, T., Fransson, I., Andersen, P.M., Dahlman-Wright, K., Fiebig, A., Sistonen, P., Savontaus, M.L., Schreiber, S., Kere, J., and Lahermo, P. (2008). Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe. PLoS ONE *3*, e3519.

20. Borodulin, K., Tolonen, H., Jousilahti, P., Jula, A., Juolevi, A., Koskinen, S., Kuulasmaa, K., Laatikainen, T., Männistö, S., Peltonen, M., et al. (2017). Cohort profile: The National FINRISK Study. Int. J. Epidemiol. *47*, 696–696i.

21. Ripatti, P., Rämö, J.T., Söderlund, S., Surakka, I., Matikainen, N., Pirinen, M., Pajukanta, P., Sarin, A.P., Service, S.K., Laurila, P.P., et al. (2016). The contribution of GWAS loci in familial dyslipidemias. PLoS Genet. *12*, e1006078.

22. Price, A.L., Weale, M.E., Patterson, N., Myers, S.R., Need, A.C., Shianna, K.V., Ge, D., Rotter, J.I., Torres, E., Taylor, K.D., et al. (2008). Long-range LD can confound genome scans in admixed populations. Am. J. Hum. Genet. *83*, 132–135, author reply 135–139.

23. Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., et al. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. *47*, 1121–1130.

24. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al.; RACI consortium; and GARNET consortium (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376–381.

25. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al.; International Multiple Sclerosis Genetics Consortium; and International IBD Genetics Consortium (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986.

26. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

27. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al.; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; and International Endogene Consortium (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature *518*, 197–206.

28. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Mägi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., et al.; ADIPOGen Consortium; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GEFOS Consortium; GENIE Consortium; GLGC; ICBP; International Endogene Consortium; LifeLines Cohort Study; MAGIC Investigators; MuTHER Consortium; PAGE Consortium; and ReproGen Consortium (2015). New genetic loci link adipose and insulin biology to body fat distribution. Nature *518*, 187–196.

29. Churchhouse, C., and Neale, B.M. (2017). Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank.http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank.

30. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al.; Electronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; and LifeLines Cohort Study (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173–1186.

31. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

32. Loh, P.R., Tucker, G., Bulik-Sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat. Genet. *47*, 284–290.

33. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: Fast and efficient meta-analysis of genome-wide association scans. Bioinformatics *26*, 2190–2191.

34. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.

35. Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. PLoS Genet. *8*, e1002453.

36. R Core Development Team (2018). A language and environment for statistical computing (R Foundation for Statistical Computing). https://www.R-project.org/.

37. Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., Boyle, E.A., Zhang, X., Racimo, F., Pritchard, J.K., and Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. eLife *8*, e39725.

38. Curtis, D. (2018). Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. Psychiatr. Genet. *28*, 85–89.

39. Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C., Chiang, C.W.K., Hirschhorn, H.N., Daly, M.J., Patterson, N., et al. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. eLife *8*, e39702.

40. McEvoy, B.P., and Visscher, P.M. (2009). Genetics of human height. Econ. Hum. Biol. *7*, 294–306.

41. Silventoinen, K., Sammalisto, S., Perola, M., Boomsma, D.I., Cornes, B.K., Davis, C., Dunkel, L., De Lange, M., Harris, J.R., Hjelmborg, J.V., et al. (2003). Heritability of adult body height: A comparative study of twin cohorts in eight countries. Twin Res. *6*, 399–408.

42. Freedman, M.L., Reich, D., Penney, K.L., McDonald, G.J., Mignault, A.A., Patterson, N., Gabriel, S.B., Topol, E.J., Smoller, J.W., Pato, C.N., et al. (2004). Assessing the impact of population stratification on genetic association studies. Nat. Genet. *36*, 388–393.

43. Marchini, J., Cardon, L.R., Phillips, M.S., and Donnelly, P. (2004). The effects of human population structure on large genetic association studies. Nat. Genet. *36*, 512–517.

44. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature *467*, 832–838.

45. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M.; and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.

46. Yengo, L., Yang, J., and Visscher, P.M. (2018). Expectation of the intercept from bivariate LD score regression in the presence of population stratification. bioRxiv. https://doi.org/10.1101/310565.

47. Scutari, M., Mackay, I., and Balding, D. (2016). Using genetic distance to infer the accuracy of genomic prediction. PLoS Genet. *12*, e1006288.

48. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2018). Clinical use of current polygenic risk scores may exacerbate health disparities. Nat. Genet. *51*, 584–591.

49. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A.; and 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

50. Lehtinen, P., Pasanen, K., Kolho, K.L., and Auvinen, A. (2016). Incidence of pediatric inflammatory bowel disease in Finland: An environmental study. J. Pediatr. Gastroenterol. Nutr. *63*, 65–70.

51. Jussila, A., Virta, L.J., Salomaa, V., Mäki, J., Jula, A., and Färkkilä, M.A. (2013). High and increasing prevalence of inflammatory bowel disease in Finland with a clear North-South difference. J. Crohn's Colitis *7*, e256–e262.

52. Lehtinen, V., Joukamaa, M., Lahtela, K., Raitasalo, R., Jyrkinen, E., Maatela, J., and Aromaa, A. (1990). Prevalence of mental disorders among adults in Finland: Basic results from the Mini Finland Health Survey. Acta Psychiatr. Scand. *81*, 418–425.

53. Hovatta, I., Terwilliger, J.D., Lichtermann, D., Mäkikyrö, T., Suvisaari, J., Peltonen, L., and Lönnqvist, J. (1997). Schizophrenia in the genetic isolate of Finland. Am. J. Med. Genet. *74*, 353–360.

54. Haukka, J., Suvisaari, J., Varilo, T., and Lönnqvist, J. (2001). Regional variation in the incidence of schizophrenia in Finland: A study of birth cohorts born from 1950 to 1969. Psychol. Med. *31*, 1045–1053.

55. Perälä, J., Saarni, S.I., Ostamo, A., Pirkola, S., Haukka, J., Härkänen, T., Koskinen, S., Lönnqvist, J., and Suvisaari, J. (2008). Geographic variation and sociodemographic characteristics of psychotic disorders in Finland. Schizophr. Res. *106*, 337–347.

56. Pietiläinen, O. (2014). Rare genomic deletions underlying schizophrenia and related neurodevelopmental disorders (University of Helsinki), PhD thesis.

57. Kurki, M.I., Saarentaus, E., Pietilainen, O., Gormley, P., Lal, D., Kerminen, S., Torniainen-Holm, M., Hamalainen, E., Rahikkala, E., Keski-Filppula, R., et al. (2019). Contribution of rare and common variants to intellectual disability in a sub-isolate of Northern Finland. Nat. Commun. *10*, 410.

58. Kaipiainen-Seppänen, O., Aho, K., and Nikkarinen, M. (2001). Regional differences in the incidence of rheumatoid arthritis in Finland in 1995. Ann. Rheum. Dis. *60*, 128–132.

Supplemental Data

Geographic Variation and Bias in the Polygenic Scores

of Complex Diseases and Traits in Finland

Sini Kerminen, Alicia R. Martin, Jukka Koskela, Sanni E. Ruotsalainen, Aki S. Havulinna, Ida Surakka, Aarno Palotie, Markus Perola, Veikko Salomaa, Mark J. Daly, Samuli Ripatti, and Matti Pirinen
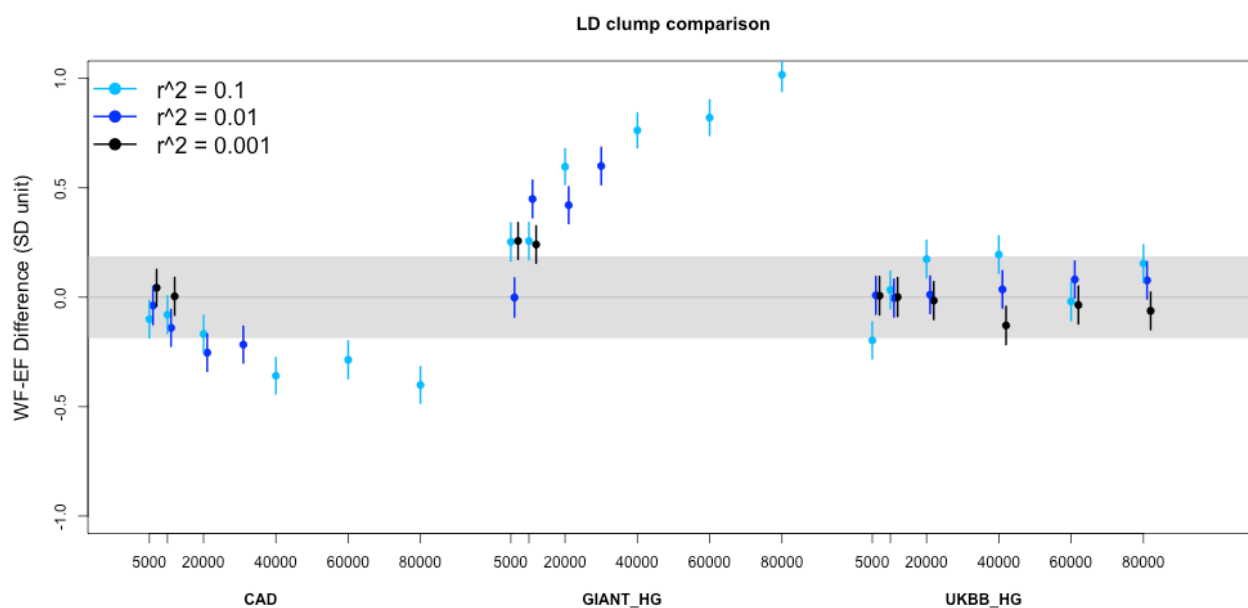
Supplemental Figures

Figure S1. Comparison of different LD-clumping thresholds.
Comparison of different r^2 thresholds on differences between Western and Eastern subpopulations in randomly chosen variants with GWAS P-value > 0.5. The solid region is the 95% probability interval under the theoretical null assumption of zero effect sizes and completely independent variants ($r^2 = 0$). Error bars refer to 95% confidence interval in t-test between subpopulations.
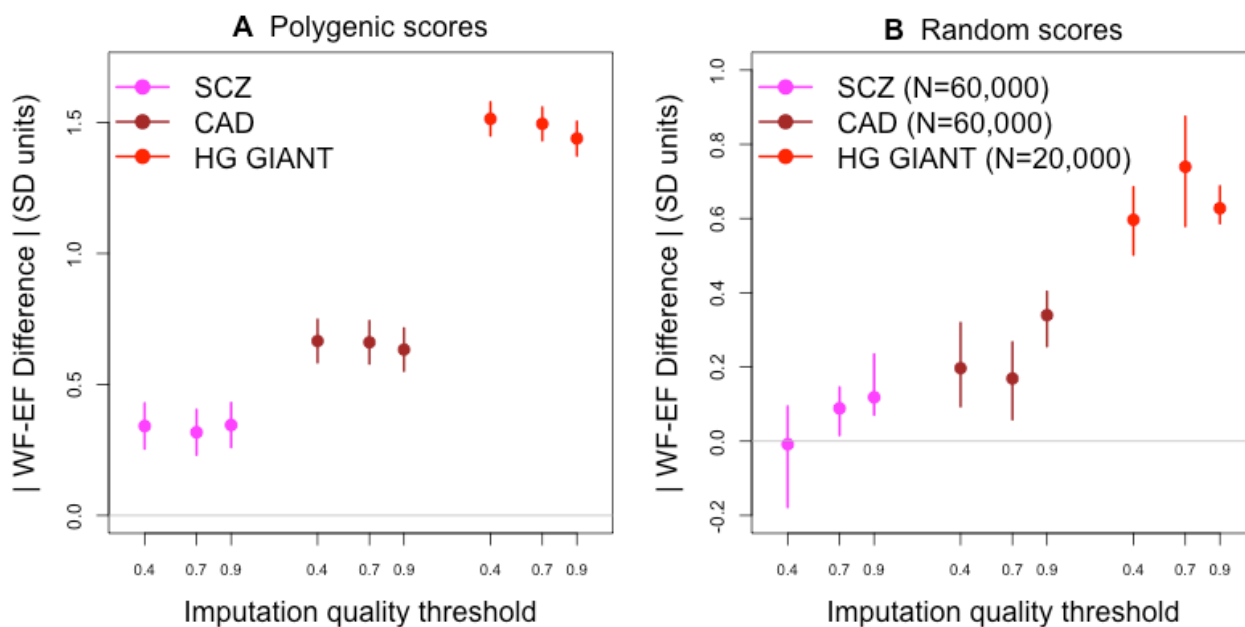

Figure S2. Comparison of different imputation quality filters.
A) comparison for trait associated polygenic scores and **B)** for random scores. Error bars in panel A) refer to 95% confidence interval in t-test between Eastern and Western subpopulations and in panel B) min-max-range from 10 randomly sampled scores with the same number of variants.
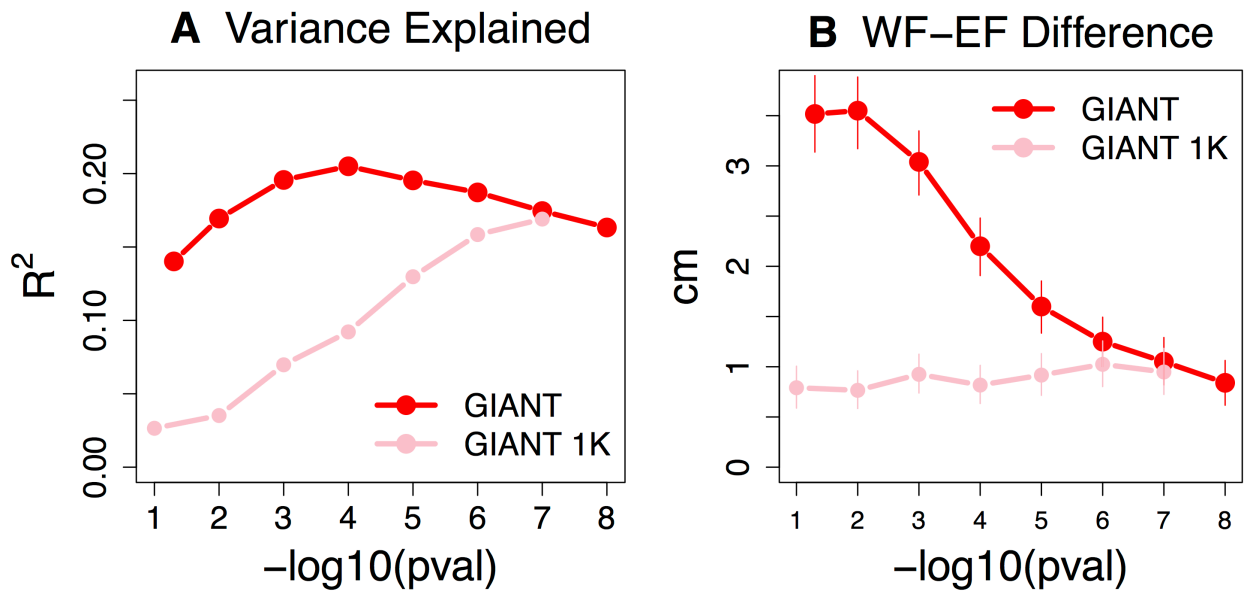
Figure S3. **Comparison of different P-value thresholds.**
A) Height variance explained by GIANT-PS and a subset of GIANT-PS with at most 1,000 randomly sampled variants, as a function of P-value threshold in GIANT data. **B)** predicted West-East difference in height by the two PS, as a function of P-value threshold in GIANT data. Variance explained is given as adjusted R². Error bars refer to 95% credible interval.
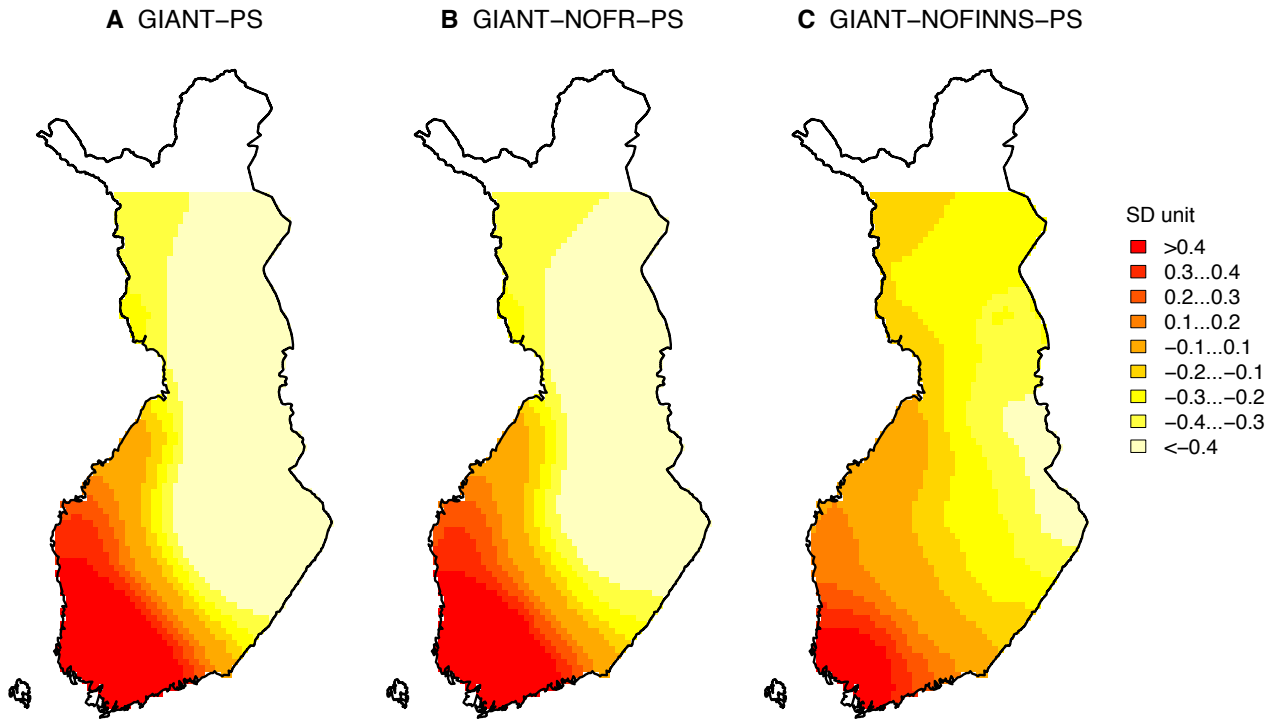
Figure S4. **Distribution of polygenic scores (PS) for HG**.

PS are based on **A)** the original GIANT consortium meta-analysis, **B)** GIANT meta-analysis without cohorts including samples from the National FINRISK Study and **C)** GIANT meta-analysis without any Finnish samples.
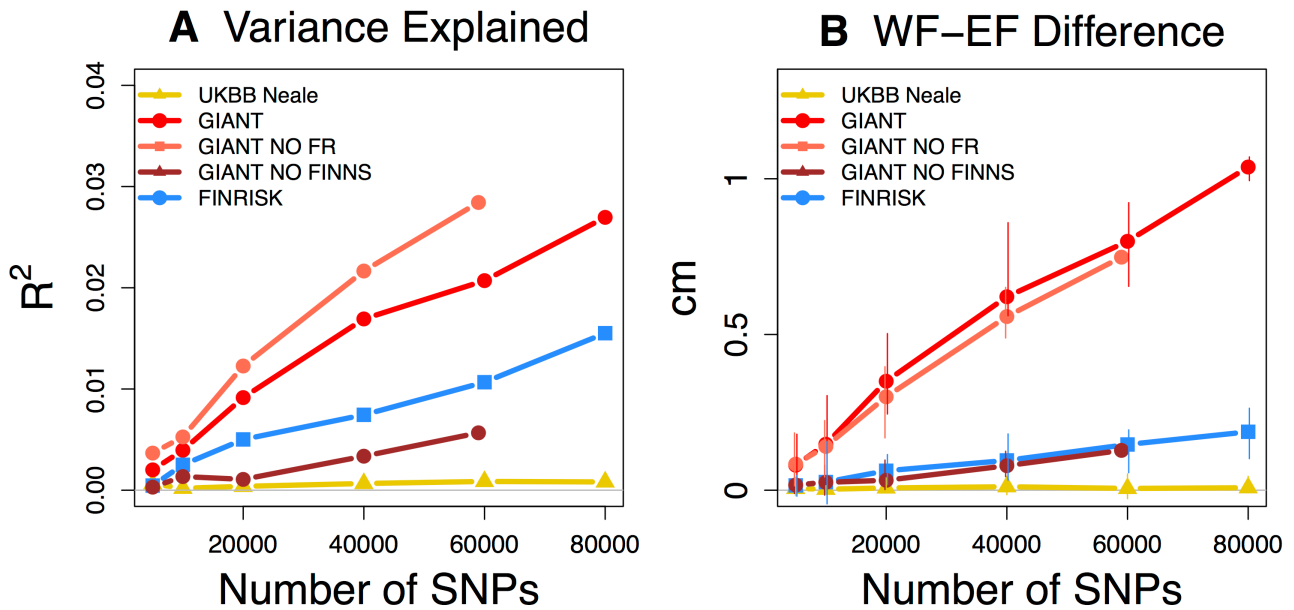

Figure S5. Comparison of PS for HG based on random SNPs.

A) Height variance explained by PS and **B)** predicted East-West difference in Height by PS, as a function of the number of independent variants in PS when all variants have P-value > 0.5 in GWAS. Variance explained is given as adjusted R^2. Both R^2 and WE difference is based on the mean of 10 random scores and error bars in panel B refer to a min-max-range of the 10 PS. For GIANT NO FR and GIANT NO FINNS we generated random PS including 59,000 variants instead of 60,000 as in other PS because of the lack of independent variants.
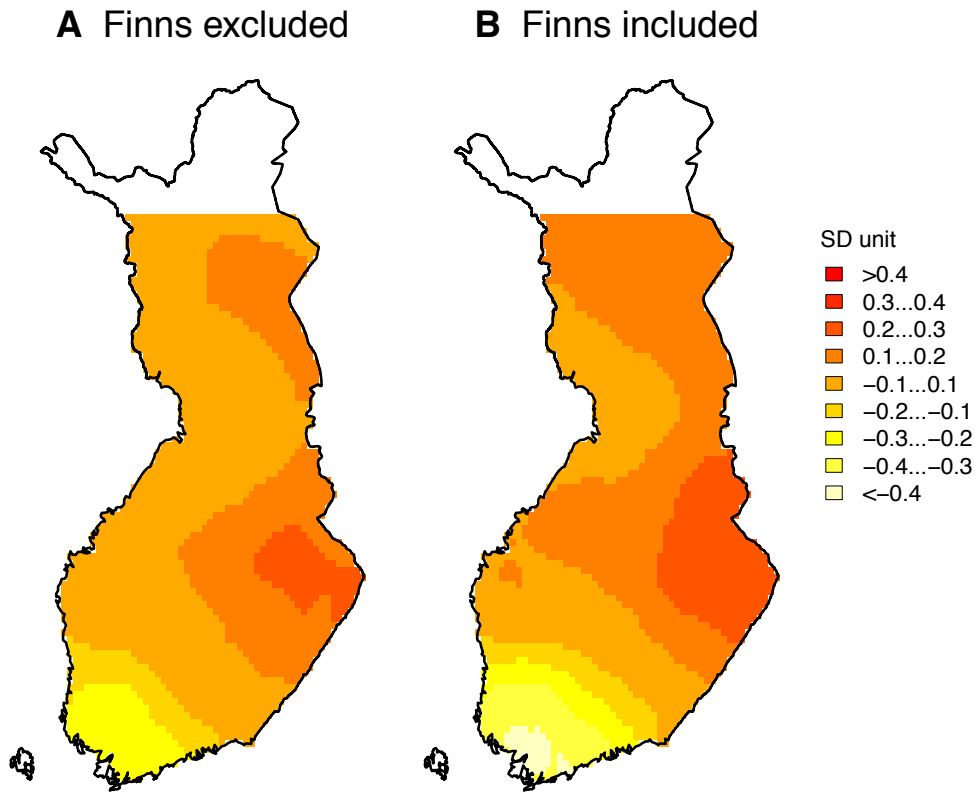
A Finns excluded **B** Finns included

SD unit
- >0.4
- 0.3...0.4
- 0.2...0.3
- 0.1...0.2
- −0.1...0.1
- −0.2...−0.1
- −0.3...−0.2
- −0.4...−0.3
- <−0.4

Figure S6. Distribution of polygenic scores for schizophrenia.
PS are based on GWAS **A)** excluding Finnish samples and **B)** including Finnish samples.
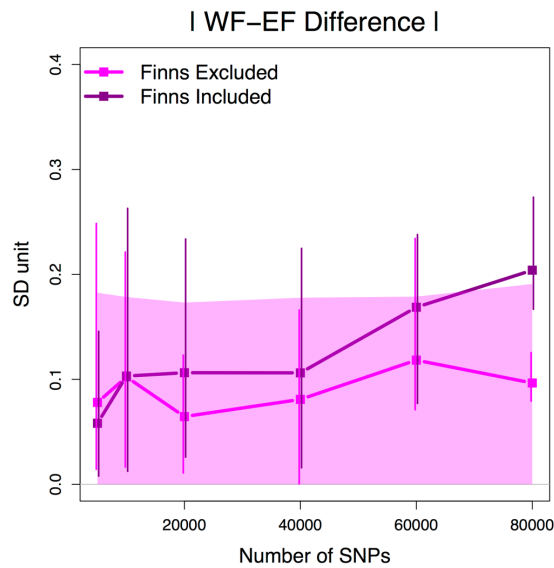

Figure S7. PS differences between Eastern and Western subpopulations using different numbers of independent variants randomly chosen with GWAS P-value > 0.5.
Results for PS built based on SCZ GWAS with excluding (magenta) and including (dark violet) Finnish samples. WE difference is based on the mean of 10 random scores and error bars refer to a min-max-range of the 10 PS.
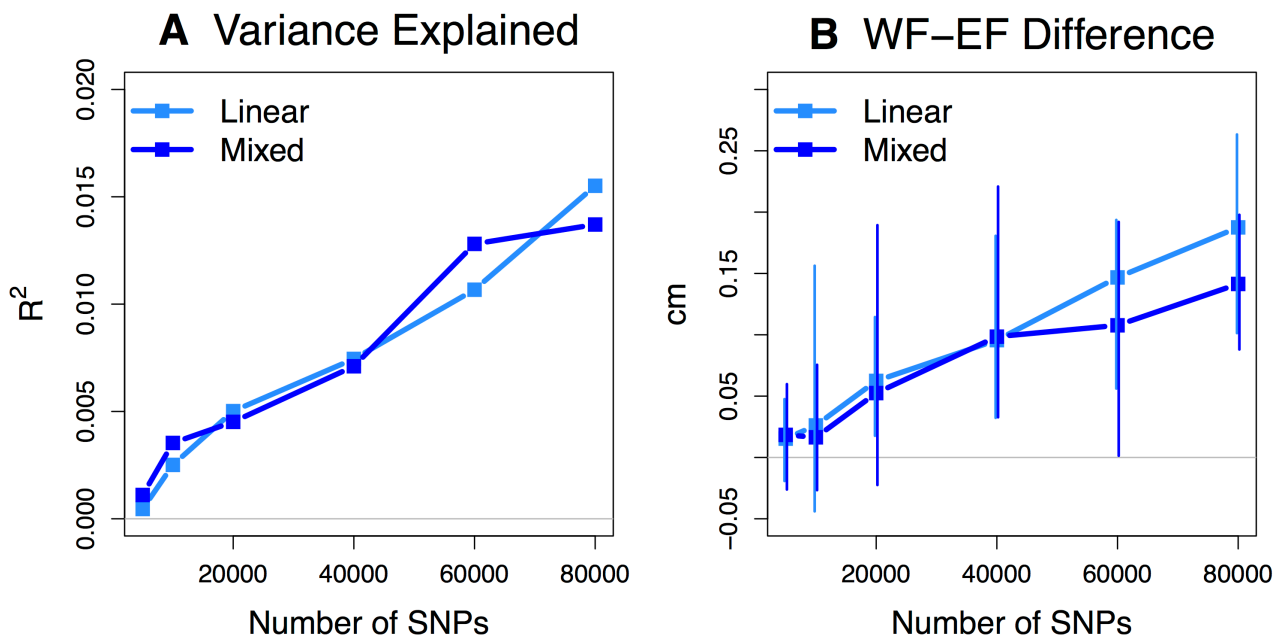
Figure S8. A comparison between random PS from the height GWAS on the FINRISK cohort ran with standard linear model or linear mixed model.
A) Height variance explained by PS and **B)** predicted West-East difference in height by PS, as a function of the number of independent variants in PS. Random PS were built based on variants with P-value > 0.5 in GWAS. Variance explained is given as adjusted R^2. Both R^2 and WF-EF difference is based on the mean of 10 random scores and error bars in panel B refer to a min-max-range of the 10 PS.
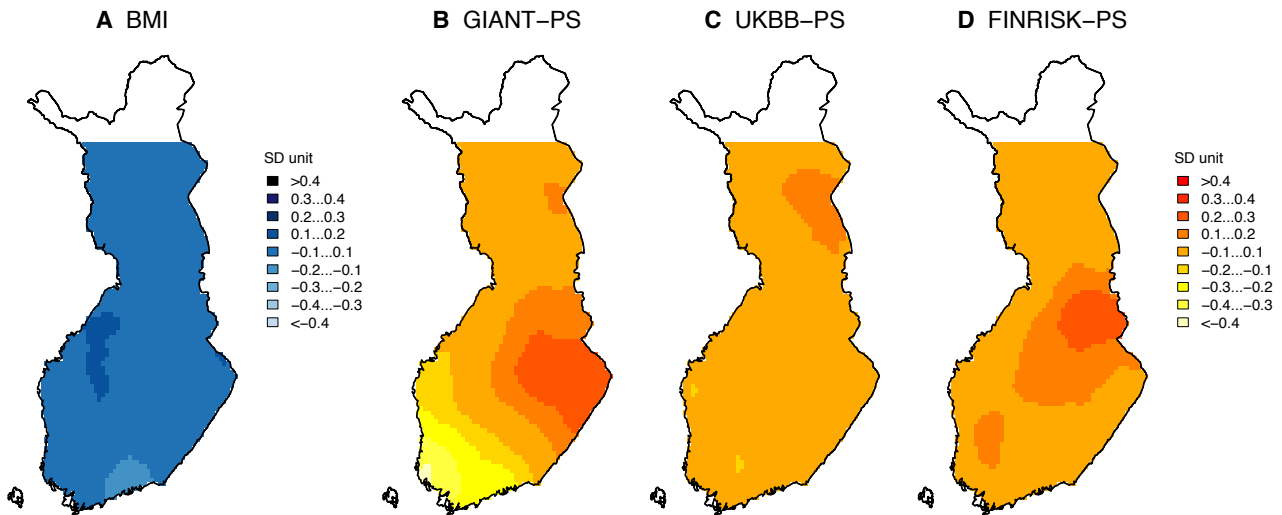
Figure S9. Geographic distributions for BMI.
A) Distribution of sex, age and age[2] adjusted BMI and polygenic score (PS) distributions of **B)** GIANT-PS,
C) UKBB-PS and **D)** FINRISK-PS for BMI in Finland. The values are in standard deviation unit. The
observed BMI does not show differences between subpopulations (95% CI: -0.12, 0.59).

Figure S10. Comparison of PS constructed from BMI-associated versus random SNPs.
Top row: **A)** BMI variance explained by PS and **B)** predicted East-West difference in BMI by PS, as a
function of P-value threshold in GWAS data.
Bottom row: **C)** BMI variance explained by PS and **D)** predicted East-West difference in BMI by PS, as a
function of the number of independent variants in PS when all variants have P-value > 0.5 in GWAS.
Variance explained is given as adjusted R^2. Both R^2 and WF-EF difference is based on the mean of 10
random scores and error bars in panel D refer to a min-max-range of the 10 PS.
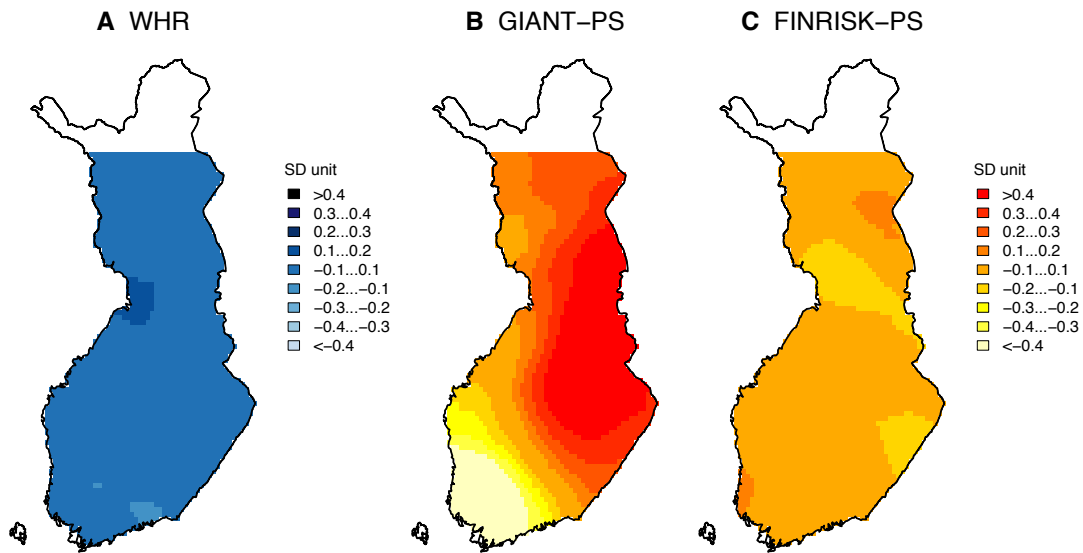
Figure S11. Geographic distributions for WHR.
A) Distribution of sex, age, age² and BMI adjusted WHR and polygenic score (PS) distributions of **B)**
GIANT-PS, **C)** FINRISK-PS for WHR (adjusted for BMI) in Finland. The values are in standard deviation
unit. The observed WHR (adjusted for BMI) does not show differences between subpopulations (95% CI: −
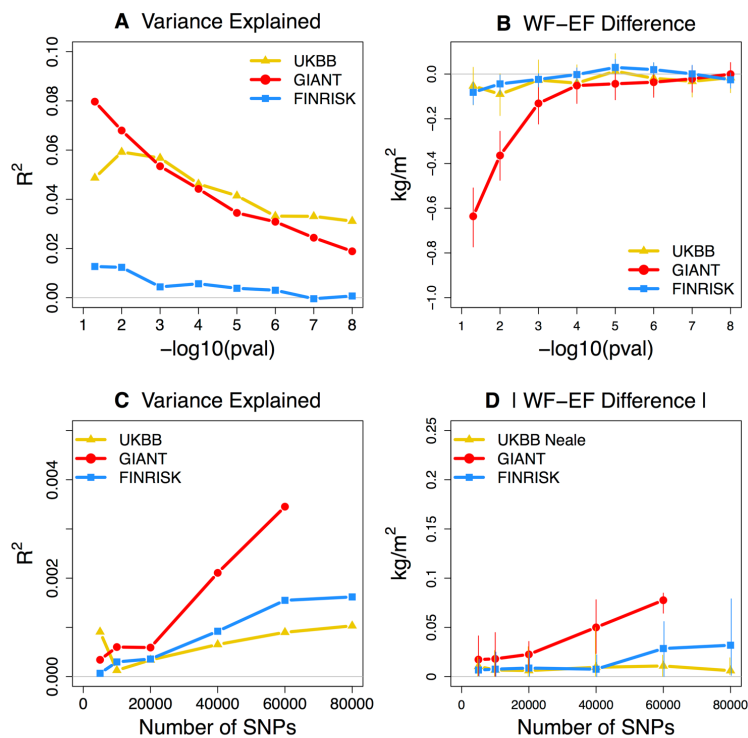0.0009, 0.0072).
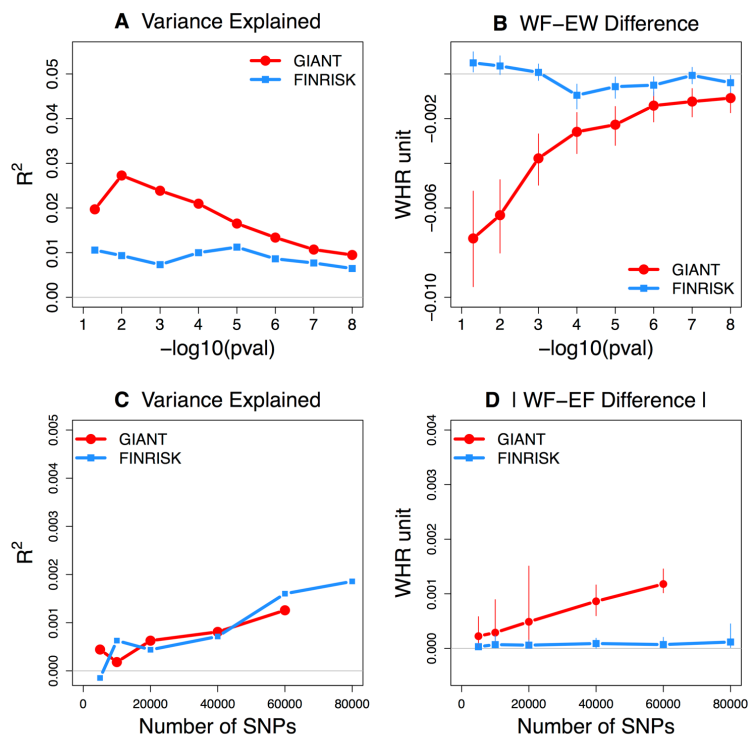
Figure S12. **Comparison of PS constructed from WHR-associated versus random SNPs.**
Top row: **A)** WHR variance explained by PS and **B)** predicted East-West difference in WHR (adjusted for
BMI) by PS, as a function of P-value threshold in GWAS data.
Bottom row: **C)** WHR variance explained by PS and **D)** predicted East-West difference in WHR (adjusted
for BMI) by PS, as a function of the number of independent variants in PS when all variants have P-value >
0.5 in GWAS. 95% CI is so small for FINRISK that it is not visible in the figure. Variance explained is
given as adjusted R². Both R² and WF-EF difference is based on the mean of 10 random scores and error bars
in panel D refer to a min-max-range of the 10 PS.

Supplemental Tables

	SNPs	Latitude		Longitude		WF-EF Difference (95% CI)
		Estimate	P-val	Estimate	P-val	
CAD	19,597	0.071	2.0e-6	0.11	4.2e-40	-0.63 (-0,71, -0.55)
RA	32,736	0.10	4.4e-11	0.11	2.2e-41	-0.63 (-0.71, -0.55)
CD	21,771	7.7e-3	6.2e-1	-0.01	1.9e-1	0.10 (0.01, 0.19)
UC	23,513	0.07	2.6e-5	0.04	2.6e-7	-0.26 (-0.35, -0.18)
SCZ	30,311	0.06	3.1e-4	0.07	3.2e-17	-0.35 (-0.43, -0.26)
BMI	12,742	0.09	2.2e-8	0.09	8.1e-31	-0.53 (-0.61, -0.44)
WHR	13,727	0.24	3.5e-60	0.18	4.0e-125	-1.16 (-1.23, -1.09)
HG	27,066	-0.35	2.7e-135	-0.27	2.8e-320	1.51 (1.45, 1.58)

Table S1. **Results from the standard linear model without accounting for genetic relatedness.**
SNPs=number of variants in polygenic score (PS). East-West difference in PS is given in standard deviation unit of PS.

	R^2			WF-EF difference (cm)		
	All (n=2,373)	East (n=1601)	West (n=772)	All	East	West
GIANT-PS	0.1402	0.1771	0.1727	3.5 (3.1, 3.9)	6.4 (5.7, 7.1)	4.7 (3.9,5.4)
UKBB-PS	0.2229	0.2126	0.2241	0.6 (0.4, 0.9)	0.6 (0.4, 0.9)	0.6 (0.4,0.9)
FINRISK-PS	0.1542	0.1581	0.104	1.4 (1.1, 1.6)	1.3 (1.1, 1.6)	1.3 (1.0, 1.6)

Table S2. Linear regression results for explaining HG with PS.
Model was applied either to all samples or separately to the eastern or western subpopulation.

	Correlation with PC1	R²		WF-EF difference (95% CI; SD units)	
		Before PC1 adjustment	After PC1 adjustment	Before PC1 adjustment	After PC1 adjustment
GIANT-PS	-0.798	0.1402	0.1916	1.51 (1.45, 1.58)	0.05 (-0.03, 0.14)
GIANT-NOFINNS-PS	-0.354	0.1721	0.1527	0.70 (0.62, 0.79)	0.04 (-0.05, 0.13)
UKBB-PS	-0.103	0.2229	0.2118	0.23 (0.14, 0.32)	0.03 (-0.06, 0.12)
FINRISK-PS	-0.304	0.1542	0.1352	0.59 (0.51, 0.67)	-0.01 (-0.09, 0.07)

Table S3. Correlation between PS and PC1 and HG variance explained (R²) by PS before and after adjustment for PC1.

	SNPs	Latitude		Longitude		WF-EF Difference (95% CI)
		Estimate	P-val	Estimate	P-val	
Finns excluded	30,311	0.04	8.7e-2	0.04	4.0e-3*	-0.35 (-0.43, -0.26)
Finns included	30,760	0.10	3.1e-6*	0.05	2.6e-4*	-0.41 (-0.49, -0.32)

Table S4. Comparison of Finns-excluded and Finns-included GWAS in schizophrenia.
Results are shown for the linear model similarly to Table 1 in the main text. SNPs=number of variants in polygenic score (PS). Difference in PS between EF and WF subpopulations is given in standard deviation unit of PS.

	FINRISK		UK Biobank	
	R²	WF-EF difference (95% CI, cm)	R²	WF-EF difference (95% CI, cm)
Standard linear model	15%	1.35 (1.14 – 1.58)	22%	0.64 (0.39 – 0.89)
Linear mixed model	15%	1.15 (0.94 – 1.38)	25%	0.37 (0.10 - 0.64)

Table S5. Comparison of PS for HG using GWAS results based on standard linear model adjusted for 10 principal components or linear mixed model in FINRISK and UK Biobank data.

| | SNPs | R^2 | Latitude | | Longitude | | WF-EF Difference (95% CI) | Correlation with PC1 | R^2 | |
			Estimate	P-val	Estimate	P-val			Before PC1 adjustment	After PC1 adjustment
GIANT-PS	12,742	8.0 %	0.03	0.094	0.04	1.8e-3	-0.64 (-0.77, -0.51)	0.286	8.0 %	7.8 %
UKBB-PS	75,979	4.9 %	0.01	0.60	0.01	0.67	-0.05 (-0.14, 0.03)	0.048	4.9 %	4.8 %
FINRISK-PS	44,920	1.3 %	-0.01	0.80	-0.002	0.90	-0.08 (-0.14, -0.04)	0.098	1.3 %	1.2 %

Table S6. Comparison of BMI PS.
(Similar to Table 2 and Table S1 for HG.)

| | SNPs | R^2 | Latitude | | Longitude | | WF-EF Difference (95% CI) | Correlation with PC1 | R^2 | |
			Estimate	P-val	Estimate	P-val			Before PC1 adjustment	After PC1 adjustment
GIANT-PS	13,130	2.0 %	0.10	1.0e-9	0.08	4.7e-12	-0.007 (-0.01, -0.005)	0.58	2.0 %	2.2 %
FINRISK-PS	43,252	1.1 %	-0.11	0.63	-0.03	0.017	5e-4 (1e-4, 1e-3)	-0.04	1.1 %	1.1 %

Table S7. Comparison of WHR PS.
(Similar to Table 2 and Table S1 for HG.)

SUPPLEMENTAL TEXT S1: DISTRIBUTION OF POLYGENIC SCORE DIFFERENCE

Consider M independent variants and assume that standard error of effect size estimate of variant k is s_k. For both quantitative traits and diseases, $s_k \approx c \cdot (2f_k(1-f_k))^{-\frac{1}{2}}$, where f_k is the minor allele frequency of variant k and c is a constant that depends on the sample size and observed phenotypic variance in the GWAS data but is independent of k.

Assume that true effects are zero at these variants whence $\widehat{\beta}_k \sim \mathcal{N}(0, s_k^2)$. Denote the z-score by $\widehat{z}_k = \widehat{\beta}_k / s_k \sim \mathcal{N}(0,1)$.

The polygenic score for individual i is

$$p_i = \sum_{k=1}^{M} g_{ik}\widehat{\beta}_k = \sum_{k=1}^{M} g_{ik}s_k\widehat{z}_k,$$

where g_{ik} is the effect allele dosage of i at variant k. By mean-centering the score, we may assume that the allele dosages were mean-centered at every variant.

Since effect estimates and genotypes are independent, $\mathrm{E}(p_i) = 0$ and

$$\mathrm{Var}(p_i) = \sum_{k=1}^{M} \mathrm{E}(g_{ik}^2 s_k^2 \widehat{z}_k^2) = \sum_{k=1}^{M} 2f_k(1-f_k) \cdot s_k^2 \cdot 1 = Mc^2.$$

Hence, the standardized score is $p_i^* = p_i/(\sqrt{M}c)$.

Consider the difference Δ between scores of a Western and an Eastern individual:

$$\Delta = p_W^* - p_E^* = \sum_{k=1}^{M} \frac{(g_{Wk} - g_{Ek})}{\sqrt{M}c} s_k \widehat{z}_k = \sum_{k=1}^{M} \frac{(g_{Wk} - g_{Ek})}{\sqrt{2Mf_k(1-f_k)}} \widehat{z}_k.$$

$\mathrm{E}(\Delta) = 0$ because $\mathrm{E}(\widehat{z}_k) = 0$ and

$$\mathrm{Var}(\Delta) = \sum_{k=1}^{M} \frac{\mathrm{Var}((g_{Wk} - g_{Ek})\widehat{z}_k)}{2Mf_k(1-f_k)} = \mathrm{Var}\left(\frac{(g_{Wk} - g_{Ek})\widehat{z}_k}{\sqrt{2f_k(1-f_k)}} \right).$$

Variance of Δ depends only on the second moment of the distribution of standardized dosage difference between West and East. In particular, variance of Δ is independent of M and sample size and phenotypic variance of the GWAS from which the effect size estimates come from. This explains why the 95% regions of Δ appear approximately constant across all traits and diseases considered.

Supplemental Text S2

To further test the possible bias in GIANT effect estimates, we took the overlapping HG variants between the GIANT-PS and the UKBB data (overlap was 26,853 out of 27,066 variants) and made a new GIANT-UKBB-PS (i.e. using GIANT variants but UKBB effects). We observed that this GIANT-UKBB-PS explained 26% of the height variance in the target sample and, contrary to GIANT-PS, its predictive power was not masked by PC1 (R^2 dropped from 26% to 24% after regressing out PC1 from GIANT-UKBB-PS). Multiple regression unequivocally showed that GIANT-PS was not predictive of HG (P = 0.25) when GIANT-UKBB-PS was simultaneously included in the model (P = 3e-84). These observations confirm that while the variants in GIANT-PS are useful for predicting HG in Finnish samples, their estimates reported in GIANT are unrealistically strongly associated with the population structure in Finland and that the UKBB effect size estimates considerably improve the predictive power of the corresponding PS. However, we also observed that, despite more accurate effect estimates, the GIANT-UKBB-PS predicted surprisingly large WF-EF difference (2.5 cm [2.0, 2.8]). This suggests that it is not only the bias in effect estimates that drives the geographic difference but also the choice of the variants.