Within-sibship GWAS improve estimates of direct genetic effects

Laurence J Howe[1,2 *], Michel G Nivard[3], Tim T Morris[1,2], Ailin F Hansen[4], Humaira Rasheed[4], Yoonsu Cho[1,2], Geetha Chittoor[5], Penelope A Lind[6,7,8], Teemu Palviainen[9], Matthijs D van der Zee[3], Rosa Cheesman[10,11], Massimo Mangino[12,13], Yunzhang Wang[14], Shuai Li[15,16,17], Lucija Klaric[18], Scott M Ratliff[19], Lawrence F Bielak[19], Marianne Nygaard[20,21], Chandra A Reynolds[22], Jared V Balbona[23,24], Christopher R Bauer[25,26], Dorret I Boomsma[3,27], Aris Baras[28], Archie Campbell[29], Harry Campbell[30], Zhengming Chen[31,32], Paraskevi Christofidou[12], Christina C Dahm[33], Deepika R Dokuru[23,24], Luke M Evans[24,34], Eco JC de Geus[3,35], Sudheer Giddaluru[36,37], Scott D Gordon[38], K. Paige Harden[39], Alexandra Havdahl[40,41], W. David Hill[42,43], Shona M Kerr[18], Yongkang Kim[24], Hyeokmoon Kweon[44], Antti Latvala[9,45], Liming Li[46], Kuang Lin[31], Pekka Martikainen[47,48,49], Patrik KE Magnusson[14], Melinda C Mills[50], Deborah A Lawlor[1,2,51], John D Overton[27], Nancy L Pedersen[14], David J Porteous[25], Jeffrey Reid[28], Karri Silventoinen[47], Melissa C Southey[17,52,53], Travis T Mallard[39], Elliot M Tucker-Drob[39], Margaret J Wright[54], Social Science Genetic Association Consortium, Within Family Consortium, John K Hewitt[23,24], Matthew C Keller[23,24], Michael C Stallings[23,24], Kaare Christensen[20,21,55], Sharon LR Kardia[19], Patricia A Peyser[19], Jennifer A Smith[19,56], James F Wilson[18,30], John L Hopper[15], Sara Hägg[14], Tim D Spector[12], Jean-Baptiste Pingault[11,57], Robert Plomin[11], Meike Bartels[3], Nicholas G Martin[38], Anne E Justice[5], Iona Y Millwood[31,32], Kristian Hveem[4,58], Øyvind Naess[36,37], Cristen J Willer[4,59,60], Bjørn Olav Åsvold[4,58,61], Philipp D Koellinger[44,62], Jaakko Kaprio[9], Sarah E Medland[6,8,63], Robin G Walters[31,32], Daniel J Benjamin[64,65,66], Patrick Turley[67,68], David M Evans[1,69], George Davey Smith[1,2], Caroline Hayward[18], Ben Brumpton[1,4,58,#], Gibran Hemani[1,2,#], Neil M Davies[1,2,4,#]

[1] Medical Research Council Integrative Epidemiology Unit, University of Bristol, BS8 2BN, United Kingdom.
[2] Population Health Sciences, Bristol Medical School, University of Bristol, Barley House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom.
[3] Department of Biological Psychology, Netherlands Twin Registry, Vrije Universiteit, Van der Boechorststraat 7, 1081 BT Amsterdam, Netherlands.
[4] K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway.
[5] Department of Population Health Sciences, Geisinger Health, 100 N. Academy Ave., Danville, PA 17822.
[6] Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia.
[7] School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia.
[8] Faculty of Medicine, University of Queensland, Brisbane, Australia.
[9] Institute for Molecular Medicine FIMM, University of Helsinki, Helsinki, Finland.
[10] PROMENTA Research Center, Department of Psychology, University of Oslo, Oslo, Norway.
[11] Social Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.
[12] Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK.
[13] NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK.
[14] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
[15] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia.
[16] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom.
[17] Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia.
[18] MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, Scotland.

[19] Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA.
[20] The Danish Twin Registry, Department of Public Health, University of Southern Denmark, Odense, Denmark.
[21] Department of Clinical Genetics, Odense University Hospital, Odense, Denmark.
[22] Department of Psychology, University of California – Riverside, Riverside, CA, USA.
[23] Department of Psychology & Neuroscience, University of Colorado at Boulder, Boulder, CO, 80309.
[24] Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, CO, 80309.
[25] BioMarin Pharmaceutical, Novato, CA.
[26] Biomedical and Translational Informatics, Geisinger Health, 100 N. Academy Ave., Danville, PA 17822.
[27] Amsterdam Public Health (APH) and Amsterdam Reproduction and Development (AR&D).
[28] Regeneron Genetics Center, Tarrytown, NY, USA.
[29] Centre for Genomic and Experimental Medicine, Institute of Genetics & Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh.
[30] Centre for Global Health Research, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland.
[31] Nuffield Department of Population Health, University of Oxford, Oxford, UK.
[32] MRC Population Health Research Unit, University of Oxford, Oxford, UK.
[33] Department of Public Health, Aarhus University, Bartholins Alle 2, 8000 Aarhus, Denmark.
[34] Department of Ecology & Evolutionary Biology, University of Colorado at Boulder, Boulder, CO, 80309.
[35] Amsterdam Public Health research institute, Amsterdam UMC, The Netherlands.
[36] Institute of Health and Society, University of Oslo.
[37] Norwegian Institute of Public Health.
[38] Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia.
[39] Department of Psychology and Population Research Center, University of Texas at Austin.
[40] Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway.
[41] Deparment of Mental Disorders, Norwegian Institute of Public Health, Oslo, Norway.
[42] Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.
[43] Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.
[44] Department of Economics, School of Business and Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.
[45] Institute of Criminology and Legal Policy, Faculty of Social Sciences, University of Helsinki, Helsinki, Finland.
[46] Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China.
[47] Population Research Unit, Faculty of Social Sciences, University of Helsinki, P.O. Box 18 (Unioninkatu 35), FIN-00014 University of Helsinki, Finland.
[48] The Max Planck Institute for Demographic Research, Germany.
[49] Department of Public Health Sciences, Stockholm University, Sweden.
[50] Leverhulme Centre for Demographic Science, University of Oxford, Oxford OX1 1JD.
[51] Bristol NIHR Biomedical Research Centre, UK.
[52] Department of Clinical Pathology, Melbourne Medicial School, The University of Melbourne, VIC, 3010, Australia.
[53] Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia.
[54] Queensland Brain Institute, The University of Queensland, Brisbane, Australia.
[55] Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark.
[56] Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104, USA.
[57] Department of Clinical, Educational and Health Psychology, University College London, London, UK.
[58] HUNT Research Center, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Levanger, Norway.
[59] Department of Internal Medicine: Cardiology, University of Michigan, Ann Arbor, MI, USA.

[60] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA.
[61] Department of Endocrinology, Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.
[62] La Follette School of Public Affairs, University of Wisconsin-Madison, 1225 Observatory Drive, Madison, WI 53706, USA.
[63] School of Psychology, University of Queensland, Brisbane, Australia.
[64] UCLA Anderson School of Management 110 Westwood Plaza Entrepreneurs Hall, Suite C515 Los Angeles, CA 90095-1481, USA.
[65] Human Genetics Department, UCLA David Geffen School of Medicine, Gonda (Goldschmied) Neuroscience and Genetics Research Center, 695 Charles E. Young Drive South, Box 708822, Los Angeles, CA 90095-7088, USA.
[66] National Bureau of Economic Research, 1050 Massachusetts Ave, Cambridge, MA 02138, USA.
[67] Center for Economic and Social Research, University of Southern California, Los Angeles, California, USA.
[68] Department of Economics, University of Southern California, Los Angeles, California, USA.
[69] University of Queensland Diamantina Institute, University of Queensland, Brisbane, 4102, Australia.
* Corresponding author: (email: laurence.howe@bristol.ac.uk)

# these authors contributed equally

Abstract

Estimates from genome-wide association studies (GWAS) represent a combination of the effect of inherited genetic variation (direct effects), demography (population stratification, assortative mating) and genetic nurture from relatives (indirect genetic effects). GWAS using family-based designs can control for demography and indirect genetic effects, but large-scale family datasets have been lacking. We combined data on 159,701 siblings from 17 cohorts to generate population (between-family) and within-sibship (within-family) estimates of genome-wide genetic associations for 25 phenotypes. We demonstrate that existing GWAS associations for height, educational attainment, smoking, depressive symptoms, age at first birth and cognitive ability overestimate direct effects. We show that estimates of SNP-heritability, genetic correlations and Mendelian randomization involving these phenotypes substantially differ when calculated using within-sibship estimates. For example, genetic correlations between educational attainment and height largely disappear. In contrast, analyses of most clinical phenotypes (e.g. LDL-cholesterol) were generally consistent between population and within-sibship models. We also report compelling evidence of polygenic adaptation on taller human height using within-sibship data. Large-scale family datasets provide new opportunities to quantify direct effects of genetic variation on human traits and diseases.

Main

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with complex phenotypes [1, 2], typically using samples of non-closely related individuals [3]. GWAS associations can be interpreted as estimates of direct individual genetic effects, i.e., the effect of inheriting a genetic variant (or a correlated variant) [4-6]. However, there is growing evidence that GWAS associations estimated from samples of unrelated individuals also capture effects of demography [7, 8] (assortative mating [9-11], population stratification [12]) and indirect genetic effects of relatives [13-19] (**Figure 1: panel A**). These non-direct sources of genetic associations are themselves of interest (e.g., for estimating parental effects [13, 18], understanding human mate choice [9-11] and genomic prediction [14, 19]) but could bias downstream analyses using GWAS summary data such as biological annotation, heritability estimation [20-22], genetic correlations [23], Mendelian randomization [7, 24, 25] and polygenic adaptation tests [26-28].

Within-family genetic association estimates, such as those obtained from samples of siblings, can provide more accurate estimates of direct genetic effects because they are unaffected by demography and indirect genetic effects of parents [7, 17, 29-32]. GWAS using siblings (within-sibship GWAS) (**Figure 1: panel B**) have been previously limited by available data, but are now feasible by combining well-established family studies with recent large biobanks that incidentally or by design contain thousands of sibships [33-36].

Here, we report findings from a within-sibship GWAS of 25 phenotypes using up to 159,701 siblings from 17 studies, the largest GWAS conducted within-sibships to date (**Figure 1: panel C**). We demonstrate that population GWAS estimates for at least 6 phenotypes (height, educational attainment, age at first birth, cognitive ability, depressive symptoms and smoking) partially reflect demography and indirect genetic effects, which affect downstream analyses such as estimates of heritability and genetic correlations. However, we find that associations with clinical phenotypes, such as lipids, are less likely to be affected. We found strong evidence of polygenic adaption on taller human height using within-sibship data. Our study illustrates the importance of collecting GWAS data from families for understanding the effects of inherited genetic variation, particularly for phenotypes sensitive to assortative mating, population stratification and indirect genetic effects.

Results

Genetic association estimates differ when accounting for demography and indirect genetic effects

For GWAS analyses we used 159,701 individuals (with one or more genotyped siblings) from 68,691 sibships in 17 studies (**Supplementary Table 1**). We used within-sibship models which, instead of estimating genotypic associations using the individual's raw genotypes and population-based samples of unrelated individuals, uses deviations of the individual's genotype from the mean genotype within the sibship (i.e. all siblings in the family present in the study). Here, the within-sibship model includes the mean sibling genotype as a covariate to capture the between-family contribution of the SNP [14]. For comparison, we also applied a standard population GWAS model that does not account for mean sibship in the same samples. Standard errors were clustered by sibship for both models. All analyses were performed in individual studies and meta-analyses were conducted across studies using summary data. Amongst the phenotypes analysed, the largest available sample sizes were for height (N = 152,350), body mass index (BMI) (N = 144,757), ever smoking (N = 125,949), systolic blood pressure (SBP) (N = 123,406)

and educational attainment (N =123,084) (**Supplementary Table 2**), we also report stratified results from non-European samples.

Previous studies have found that association estimates of height and educational attainment genetic variants are smaller in within-family models [13, 14, 37]. We aimed to investigate whether similar shrinkage is observed for other phenotypes by comparing within-sibship and population genetic association estimates for 25 phenotypes. We observed the largest within-sibship shrinkage for genetic variants associated with age at first birth (49%, 95% C.I. [25%, 74%]) and educational attainment (46%, [40%, 52%]). We also found evidence of shrinkage for depressive symptoms (39%, [5%, 73%]), ever smoking (19%, [9%, 30%]), cognitive ability (18%, [2%, 35%]) and height (10%, [8%, 12%]). In contrast, within-sibship association estimates for C-reactive protein (CRP) were larger than population estimates (-9%, [-15%, -2%]). We found limited evidence of within-sibship differences for the remaining 18 phenotypes, including BMI and SBP (**Figure 2 / Supplementary Table 3**).

We investigated possible heterogeneity in shrinkage for height and educational attainment genetic variants across individual variants and between cohorts. In the meta-analysis data, we observed minimal evidence of heterogeneity in shrinkage across individual variants for height and educational attainment, suggesting that shrinkage is largely uniform across the strongest association signals for these phenotypes. We also found limited evidence of cohort heterogeneity in shrinkages for height (heterogeneity P = 0.25) and educational attainment (P = 0.17) across the European-ancestry cohorts (**Supplementary Figures 1/2**). In contrast, there was limited evidence for shrinkage on height in China Kadoorie Biobank (shrinkage -3%; 95% C.I. [-13%, 7%]; heterogeneity with European meta-analysis P-value = 0.005) but some evidence of shrinkage on ever smoking (shrinkage = 134%; [10%, 258%]) (**Supplementary Figure 3**).

Within-sibship SNP heritability estimates

LD score regression (LDSC) can use GWAS data to estimate SNP heritability, the proportion of phenotypic variation explained by common SNPs [20, 23]. We used simulations to investigate the applicability of LDSC when using within-sibship GWAS data, finding evidence that LDSC can estimate SNP heritability using both population and within-sibship model GWAS data if effective sample sizes (based on standard errors) are used to account for differences in power between the models (**Methods**).

To evaluate the impact of controlling for demography and indirect genetic effects, we compared LDSC SNP heritability estimates based on population and within-sibship effect estimates for 25 phenotypes. Theoretically, within-sibship shrinkage in GWAS effect estimates will also lead to attenuations in within-sibship SNP heritability estimates (**Methods**). The within-sibship SNP heritability point estimate for educational attainment attenuated by 71% from the population estimate (Population h^2: 0.14, within-sibship h^2: 0.04, difference P = 1.5×10^{-20}) with attenuations also observed for cognition (Population h^2: 0.24, within-sibship h^2: 0.13, attenuation 46%; difference P = 0.012) and height (Population h^2: 0.41, within-sibship h^2: 0.34, attenuation 17%; difference P = 1.3×10^{-3}). The observed attenuations were consistent with theoretical expectation (**Supplementary Table 4**), suggesting that the lower within-sibship SNP heritability estimates are explained by association estimate shrinkage. Across the 22 additional phenotypes, population and within-sibship SNP heritability estimates were relatively consistent (**Figure 3 / Supplementary Table 5**).

Within-sibship genetic correlations with educational attainment

We used LDSC [23] to estimate cross-phenotype genome-wide genetic correlations (r_g) between educational attainment and 22 phenotypes with sufficient heritability (Population/within-sibship $h^2 > 0$). To determine the effects of demography and indirect genetic effects on r_g, we compared estimates of r_g using population and within-sibship estimates.

There was strong evidence using population estimates that educational attainment is positively correlated with height (r_g = 0.16, 95% C.I [0.10, 0.22]) and negatively correlated with BMI (r_g = -0.32, [-0.38, -0.26]) and CRP (r_g = -0.47, [-0.69, -0.25]). However, these correlations were negligible when using within-sibship estimates; height (r_g = -0.02, [-0.15, 0.10]), BMI (r_g = 0.01, [-0.17, 0.19]) and CRP (r_g = 0.04, [-0.34, 0.42]) with evidence for differences between population and within-sibship r_g estimates (height difference P = 4.0×10^{-3}, BMI difference P = 6.9×10^{-5}, CRP difference P = 0.012). These attenuations indicate that genetic correlations between educational attainment and these phenotypes from population estimates are likely to be driven by demography and indirect genetic effects (**Figure 4 / Supplementary Table 6**).

Within-sibship Mendelian randomization: effects of height and BMI

Mendelian randomization uses genetic variants as instrumental variables to assess the causal effect of exposure phenotypes on outcomes [24, 38]. Mendelian randomization was originally conceptualised in the context of parent-offspring trios where offspring inherit a random allele from each parent [24]. However, with limited family data, most Mendelian randomization studies have used data from unrelated individuals. Within-sibship Mendelian randomization (WS-MR) is largely robust against demography and indirect genetic effects that could distort estimates from non-family designs [7, 25]. Here, we used population and WS-MR to estimate the effects of height and BMI on 23 phenotypes.

WS-MR estimates for height and BMI on the 23 outcome phenotypes were largely consistent with population MR estimates for height (0%; 95% C.I. -12%, 12%) but slightly lower for BMI (-11%; 95% C.I. -20%, -1%). However, consistent with the genetic correlation analyses, we observed differences between population and WS-MR estimates of height and BMI on educational attainment. Population Mendelian randomization estimates provided strong evidence that taller height and lower BMI increase educational attainment (0.05 SD increase in education per SD taller height, 95% C.I. [0.03, 0.06]; 0.18 SD decrease in education per SD higher BMI, [0.14, 0.21]). In contrast, WS-MR estimates for these relationships were greatly attenuated (height: 0.01 SD increase [-0.01, 0.03], difference P = 2.9×10^{-3} | BMI; 0.04 SD decrease [0.00, 0.08], difference P = 2.1×10^{-7}). We also observed similar attenuation from population and WS-MR estimates for BMI on age at first birth (difference P = 6.0×10^{-4}); a phenotype highly correlated with education, and some suggestive evidence of attenuation in the WS-MR estimate of height on triglycerides (difference P = 0.04). These differences illustrate instances where population based Mendelian randomization estimates are distorted by demography and indirect genetic effects (**Table 1**).

Polygenic adaptation

Polygenic adaptation is a process via which phenotypic changes in a population over time are induced by small shifts in allele frequencies across thousands of variants. One method of testing for polygenic adaptation is to compare Singleton Density Scores (SDS), measures of natural selection over the previous 2,000 years [28], with GWAS P-values. However, this approach is sensitive to population stratification as illustrated by recent work using UK Biobank data which showed that population stratification in GWAS data likely confounded previous estimates of polygenic adaptation on height [26, 27]. Within-sibship GWAS data is particularly useful in this context as it is robust against population stratification. Here we re-calculated Spearman's rank correlation (r) between tSDS (SDS scores aligned with the phenotype increasing allele) and our population/within-sibship GWAS P-values for 25 phenotypes, with standard errors estimated using jack-knifing over blocks of genetic variants.

We found strong evidence for polygenic adaptation on taller height in the European meta-analysis GWAS using both population (r = 0.020, 95% C.I. [0.011, 0.029]) and within-sibship GWAS estimates (r = 0.011, [0.003, 0.020]) (**Supplementary Figures 4/5**). These results were supported by several sensitivity analyses; a) evidence of enrichment for positive tSDS (mean = 0.23, SE = 0.06, P < 0.001) amongst 416

putative height loci from the within-sibship meta-analysis data (**Supplementary Figure 6**), b) positive LDSC r_g between height and tSDS in the meta-analysis data (**Supplementary Table 7**) and c) evidence for polygenic adaptation on taller height when meta-analysing correlation estimates from 7 individual studies (e.g. SDS using only UK Biobank GWAS summary data) for population (r = 0.013, [0.010, 0.015]) and within-sibship (r = 0.004, [0.002, 0.007]) estimates (**Figure 5**). There was also some putative within-sibship evidence for polygenic adaptation on increased number of children (P = 0.006) and lower HDL-cholesterol (P = 0.024) (**Supplementary Figure 4**).

Discussion

These results demonstrate that GWAS results and downstream analyses of behavioural phenotypes (e.g. educational attainment, smoking behaviour) as well as some biologically proximal phenotypes (e.g. height, BMI) are likely to be affected by demography and indirect genetic effects. However, we found that most analyses involving more clinical phenotypes, such as lipids, were not strongly affected. Future studies should use data from unrelated individuals, to maximise sample size for gene discovery and polygenic prediction, and data from families to provide more accurate estimates of direct genetic effects for downstream analyses.

A key aim of GWAS is to identify direct genetic effects on phenotypes, but other sources of genetic associations can be of value for analyses. For example, knowledge of indirect genetic effects can be used to elucidate maternal effects [15, 39] or the extent to which health outcomes are mediated by family environments [13, 18]. Future family based GWAS will also enable the estimation of indirect genetic effects [6, 18, 40].

We observed minimal evidence of heterogeneity in shrinkage estimates of height and educational attainment genetic variants. This indicates that observed shrinkage is likely to be largely driven by assortative mating or indirect genetic effects since both of these tend to influence associations proportional to the direct effect (whereas population stratification is likely to have larger effects on ancestrally informative markers). The common environment terms from classical twin studies suggest that there are likely to be indirect genetic effects on educational attainment [41], cognitive ability [42] and smoking [43], but suggest that the observed shrinkage for height is likely to be a consequence of assortative mating [10, 43, 44].

Within-sibship GWAS data can be useful for validating results from larger samples of unrelated individuals. Here, we showed that population and within-sibship Mendelian randomization estimates of height and BMI were generally consistent for 23 outcome phenotypes. However, we observed differences between within-sibship and population Mendelian randomization estimates of height (on educational attainment and triglycerides) and BMI (on educational attainment and age at first birth), suggesting the Mendelian randomization assumptions do not hold for these relationships in samples of unrelated individuals. For future Mendelian randomization studies, within-sibship estimates could elucidate the potential presence of bias [7, 25].

We used non-European data from China Kadoorie Biobank to evaluate whether demography and indirect genetic effects influence GWAS analyses conducted in the Chinese population. In this sample, we found minimal evidence of shrinkage for height genetic variants but (consistent with the European meta-analysis) suggestive evidence of shrinkage for variants associated with smoking initiation. The absence of shrinkage for height suggests that demographic effects such as assortative mating may differ between populations. Larger within-family studies in non-European populations could be used to evaluate population differences in demographic and indirect effects.

We also used the within-sibship GWAS data to evaluate evidence for recent selection. A previous study reporting polygenic adaptation on height in the UK population was found to be biased by population stratification in the GIANT consortium [26-28]. Previous evidence for adaptation on height using siblings in UK Biobank was suggestive of some adaptation, but statistically inconclusive [26]. Here, using within-sibship GWAS estimates from a larger (~4-fold) sample of siblings, we found strong evidence of polygenic adaptation on increased height and some evidence of adaptation on number of children and HDL-cholesterol. We anticipate that future studies on human evolution will benefit from using large within-family datasets such as our resource.

Within-family GWAS are limited by both available family data and statistical inefficiency (homozygosity within-families). To address this issue, future population-based biobanks could recruit the partners, siblings and offspring of study participants. In contrast, conventional GWAS designs sampling unrelated individuals are likely to be the optimal approach to maximise statistical power for discovery GWAS for genetic associations. Indeed, we found that many genotype-phenotype associations from population GWAS models were also observed in within-sibship GWAS, albeit sometimes with attenuated association estimates. A notable limitation of within-sibship models is that they do not control for indirect genetic effects of siblings, i.e. effects of sibling genotypes on the shared environment. Sibling effects have been estimated to be modest compared to parental effects [6, 45] but could have impacted our GWAS estimates. Our findings are also limited to adult phenotypes. Future within-family GWAS (e.g. using parent-offspring trios) could use data from children to evaluate if childhood phenotypes are more strongly affected by indirect genetic effects.

Methods

Study participants

Eighteen cohorts contributed data to the overall study (**Supplementary Table 1**). These cohorts were selected on the basis of having at least 500 genotyped siblings (an individual with 1 or more siblings in the study sample) with at least 1 of the 25 phenotypes that were analysed in the study. Detailed information on genotype data, quality control and imputation processes are provided in the **Supplementary Materials**. Individual cohorts defined each phenotype based on suggested definitions from an analysis plan (see the **Supplementary Materials**).

GWAS analyses

GWAS analyses were performed uniformly across individual studies using automated scripts and a pre-registered analysis plan (https://github.com/LaurenceHowe/SiblingGWAS). Scripts checked strand alignment, imputation scores and allele frequencies for the genetic data as well as missingness for covariates and phenotypes. Scripts also summarised covariates and phenotypes and set phenotypes to missing for sibships if only one individual in the sibship has non-missing phenotype data. To harmonise variants for meta-analysis, genetic variants were renamed in a format including information on chromosome, base pair, and polymorphism type (SNP or INDEL). The automated pipeline restricted analyses to common genetic variants (MAF > 0.01) and removed poorly imputed variants (INFO < 0.3). Analyses were restricted to include individuals in a sibship, i.e. a group of two or more full siblings in the study. Monozygotic twins were included if they had an additional sibling in the study.

GWAS analyses involved fitting both population and within-sibship models to the same samples. The population model is synonymous with a conventional principal component adjusted model, and was fit using linear regression in R. The within-sibship model is an extension of the population model including

the family mean genotype (the mean genotype of siblings in each family) as a covariate to account for family structure with each individual's genotype centred around the family mean [7, 14]. Age, sex and up to 20 principal components (10 principal components were included in smaller studies at the discretion of study co-authors) were included as covariates in both models.

For individual j in sibship i with $n_i \geq 2$ siblings:

Population model:

$$Phenotype_{ij} \sim G_{ij} + Sex_{ij} + Age_{ij} + PC1_{ij} + .. PC20_{ij}$$

Within-sibship model:

$$Phenotype_{ij} \sim G_{ij}^{C} + G_{i}^{F} + Sex_{ij} + Age_{ij} + PC1_{ij} + .. PC20_{ij}$$

where $G_i^F = \frac{\sum_1^n G_{ik}}{n}$ and $G_{ij}^C = G_{ij} - G_i^F$

G_{ij} = genotype of sibling j in sibship i, G_i^F = mean family genotype for sibship i over n siblings, G_{ij}^C = genotype of sibling j in sibship i centred around G_i^F.

Standard errors from both estimators were clustered over families at the sibling level to account for non-random clustering of siblings within families. Note that this clustering accounts for sibling relationships but does not account for further relatedness present in each sample. For example, a sibling pair could be related to another sibling pair (i.e. two pairs of siblings who are first-cousins). We performed simulations, described below, confirming that such relatedness can lead to underestimating standard errors in the population model and has no effect on the standard errors of the within-sibship model.

GWAS models were performed in individual studies, harmonised and then meta-analysed for each phenotype using a fixed-effects model in METAL [46] with population and within-sibship data meta-analysed separately. We performed meta-analyses using only samples of European ancestry. We used data from 13,856 individuals from China Kadoorie Biobank separately in downstream analyses. The largest meta-analysis sample sizes were for height (N = 137,776), BMI (N = 130,183), ever smoking (N = 111,375), SBP (N = 108,832) and educational attainment (N = 108,510) (**Supplementary Table 2**).

Further quality control was performed prior to meta-analyses. We used phenotype-specific genotype counts (e.g. from the sample with height data in a study) to exclude variants missing in more than 10% of samples. We found some evidence that low frequency variants in small sample sizes may have inflated test statistics in regression models. We randomly selected two sets of 250 sibship (~500 individuals) in UK Biobank and performed population and within-sibship GWAS. We found high levels of test statistic inflation with hundreds (population model) or thousands (within-sibship model) of genome-wide significant hits despite the small sample size. These variants were found to be overwhelmingly low frequency; the 99th percentile MAF for the genome-wide significant variants were 3.0%/2.3% in within-sibship model and 2.9%/4.6% in the population model. Therefore, we used phenotype-specific MAFs and study-level INFO to perform additional stringent quality control on the GWAS data using the following cut-offs for INFO and MAF; fewer than 1,000 individuals (MAF < 0.1, INFO < 0.8); 1,000-3,000 individuals (MAF < 0.05, INFO < 0.5); 3,000-5,000 individuals (MAF < 0.03, INFO < 0.5); 5,000-10,000 individuals (MAF < 0.02, INFO < 0.3); and more than 10,000 individuals (MAF < 0.01, INFO < 0.3).

Meta-analysis

Phenotypes were harmonised between studies using phenotypic summary data on means and standard deviations. GWAS which did not conform to analysis plan definitions (e.g. binary instead of continuous) were excluded from meta-analyses. GWAS presented in different continuous units (e.g. not standardised) were transformed before meta-analysis by dividing association estimates and standard errors by the standard deviation of the phenotype as measured in the cohort. Meta-analyses for 25 phenotypes were performed using a fixed-effects model in METAL [46].

Within-sibship and population-based GWAS comparison

Overview

We hypothesised that the within-sibship estimates would differ compared to population-based estimates due to the exclusion of effects from demographic and familial pathways. In general, these effects have been shown to inflate (rather than shrink) population-based estimates so we estimated within-sibship shrinkage (the % difference from population to within-sibship estimates). To estimate this shrinkage, we required estimates of the associations with a phenotype from each within-sibship and population-based analyses that were not affected by winner's curse. Hence, we adopted a strategy where we used an independent reference dataset to select the variants associated with a phenotype. Using the meta-analysis results to obtain association estimates for these variants, we generated summary-based weighted scores of those association estimates in the within-sibship and population-based analyses and estimated the ratio of those scores. We used the UK Biobank dataset that excludes sibling data as the independent reference dataset.

GWAS in independent reference discovery dataset

We performed GWAS in an independent sample of UK Biobank (excluding siblings) for each phenotype using a linear mixed model as implemented in BOLT-LMM [47]. We started with a sample of 463,006 individuals of 'European' ancestry derived using in-house k-means cluster analysis performed using the first 4 principal components provided by UK Biobank with standard exclusions also removed [48]. To remove sample overlap, we then excluded the sibling sample (N = 40,276), resulting in a final sample of 422,730 individuals. To model population structure in the sample, we used 143,006 directly genotyped SNPs, obtained after filtering on MAF > 0.01; genotyping rate > 0.015; Hardy-Weinberg equilibrium p-value < 0.0001 and LD pruning to an r^2 threshold of 0.1 using PLINK v2.0 [49]. Age and sex were included in the model as covariates.

All 25 phenotypes (conforming to our phenotype definition) were available in UK Biobank data except for a continuous measure of depressive symptoms. For depressive symptoms, we performed a GWAS of binary depression which was excluded from the meta-analysis (see definition in **Supplementary Materials**). Using the BOLT-LMM UK Biobank GWAS data, we performed strict LD clumping in PLINK v2.0 [49] ($r^2 < 0.001$, physical distance threshold = 10,000 kb) to generate independent variants associated with each phenotype at genome-wide significance ($P < 5\times10^{-8}$) and at a more liberal threshold ($P < 1\times10^{-5}$).

Summary-based weighted scores

For a particular phenotype the sets of independent variants obtained from the independent UK Biobank GWAS were used to generate a summary-based weighted score using an inverse variance weighted (IVW) approach [50, 51]:

$$S = \frac{\sum_k^M \frac{w_k \beta_k}{\sigma_k^2}}{\sum_k^M \frac{w_k^2}{\sigma_k^2}}$$

with standard error

$$\sigma_S = \sqrt{\frac{1}{\sum_k^M \frac{w_k^2}{\sigma_k^2}}}$$

Here the score S represents the weighted average of the association estimates of the M variants on a phenotype, where β and σ represent the beta coefficients and standard errors from the within-sibship (W) or population-based (P) meta-analysis results. The discovery association estimates from the UK Biobank GWAS were used as weights (w). The set of M variants were determined using either the genome-wide significance (G) or the more liberal threshold (L). Hence, depending on which model is used to determine the association estimates and which set of SNPs are used, for each phenotype four scores can be calculated – $S_{P,G}$, $S_{P,L}$, $S_{W,G}$ and $S_{W,L}$.

These sets of scores were obtained for each of the 25 phenotypes with weights for binary depression used as a substitute for depressive symptoms because a suitable measure was unavailable in UK Biobank. The scores were strongly associated with the set of phenotypes in the meta-analysis data based on determining p-values from their Z-scores. The $S_{W,L}$ scores were nominally associated at $p < 0.05$ for 24 out of 25 (exceptions: number of children) of the phenotypes with the $S_{P,L}$ scores associated with all 25 phenotypes at this threshold (**Supplementary Table 8**).

Estimating shrinkage from population to within-sibship estimates

We used the within-sibship and population-based scores to calculate the average shrinkage (δ, i.e. proportion decrease) of genetic variants-phenotype associations

$$\delta_{\cdot} = 1 - \frac{S_{W,\cdot}}{S_{P,\cdot}}$$

The standard errors of δ could be estimated using the delta method as below using the standard errors of the scores and the covariance between the scores $Cov\left(S_{w,\cdot}, S_{P,\cdot}\right)$:

$$\sigma_{\delta_{\cdot}} \sim \left(\frac{S_{W,\cdot}}{S_{P,\cdot}}\right) \sqrt{\left(\frac{\sigma^2_{S_{W,\cdot}}}{S^2_{W,\cdot}} + \frac{\sigma^2_{S_{P,\cdot}}}{S^2_{P,\cdot}}\right) - \frac{2Cov\left(S_{w,\cdot}, S_{P,\cdot}\right)}{S_{W,\cdot} S_{P,\cdot}}}$$

However, we do not have an estimate of this covariance term because the two GWAS were fit in separate regression models. We therefore used the jackknife to estimate $\sigma_{\delta_{\cdot}}$. For a score of M variants, we removed each variant in turn and repeated IVW and shrinkage analyses as above, extracting the shrinkage point estimate in each of the M iterations. We then calculated $\sigma_{\delta_{\cdot}}$ as follows:

$$\sigma_{\delta_{.}} = \sqrt{\frac{M-1}{M}\sum_{k}^{M}(\sigma_{\delta_{.,k}} - \mu)^2}$$

where

$$\mu = \frac{\sum_{k}^{M}\sigma_{\delta_{.,k}}}{M}$$

As a sensitivity analysis, we investigated the effects of positive covariance between the population and within-sibship models on the shrinkage standard errors using individual-level participant data from UK Biobank. Analysing shrinkage on height, we used seemingly unrelated regression (SUR) to estimate the covariance term between the population and within-sibship estimators. We found that standard errors for shrinkage estimates decreased by around 15% when the covariance was modelled (**Supplementary Table 9**). SUR standard errors were consistent with the jackknife approach standard errors.

As the primary analysis, we reported shrinkage results using the liberal threshold (P-value < $1x10^{-5}$) with results using the genome-wide threshold (P-value < $5x10^{-8}$) reported as a sensitivity analysis. In the main text, we report the shrinkage estimates that reach nominal significance (P < 0.05). We presented shrinkage estimates in terms of % (multiplying by 100).

As a sensitivity analysis, we also presented study-level shrinkage estimates for height and educational attainment and tested for heterogeneity. These phenotypes were chosen because of previous evidence for shrinkage on these phenotypes and available data.

Heterogeneity of shrinkage across variants within a phenotype

We used results of the within-sibship and population-based meta-analyses to estimate whether shrinkage estimates were consistent across all variants within a phenotype, using an estimate of heterogeneity. As above, we only evaluated heterogeneity for height and educational attainment because of previous evidence and available data. For each variant we estimated the Wald ratio of the shrinkage estimate

$$s_k = \frac{\beta_{P,k}}{\beta_{W,k}}$$

The heterogeneity estimate was obtained as

$$Q = \sum_{k}^{M} w_k^2 (s_k - S)^2$$

where

$$w_k = \sqrt{\frac{S^2}{\sigma_{W,k}^2 + S^2\sigma_S^2}}$$

Applying LD score regression to within-sibship data

LDSC is a widely used method that can be applied to GWAS summary data to estimate heritability and genetic correlation [20, 23]. Central to the method is that LDSC can detect and control for confounding (which is not correlated with LD scores) in GWAS data such as from cryptic relatedness and population stratification. The LDSC ratio, a function of the LDSC intercept unrelated to statistical power, is a measure of the proportion of association signal that is due to confounding. Notably the LDSC ratio will not identify sources of association that are correlated with LD scores such as indirect genetic effects or assortative mating as confounding. We therefore loosely interpret the LDSC ratio as a measure of confounding as it will not identify all sources of confounding. In this work, we apply LDSC to estimate SNP heritability and genetic correlation using the population and within-sibship GWAS data, so we investigated the LDSC intercept/ratio estimates from these data.

In theory, within-sibship data should be less susceptible to confounding than population data as it more effectively controls for population stratification than including principal components. To investigate this in practice, we used LDSC to estimate confounding in meta-analysis summary data for 25 phenotypes. Summary data were harmonised using the LDSC munge_sumstats.py function. LDSC intercepts and ratios were estimated using the harmonised data and the LDSC ldsc.py function with the precomputed European LD scores from the 1000 Genomes (Phase 3) reference panel. The LDSC ratio was used for comparisons between phenotypes and studies as it is not a function of statistical power. The LDSC ratio is calculated from the intercept (i) and the mean chi squared χ^2 as follows:

$$Ratio = \frac{i - 1}{mean(\chi^2) - 1}$$

LDSC confounding estimates varied across the 25 phenotypes in the within-sibship model. Confounding estimates were modest for height (11%, 95% C.I. [7%, 14%]) and BMI (6%, [-1%, 14%]) while the estimate for educational attainment was imprecise (35%, [12%, 58%]). Across all phenotypes in the within-sibship data, the median confounding estimate was 21% (Q1-Q3: 10%, 33%) but stronger conclusions are limited by imprecise estimates (**Supplementary Table 10/ Supplementary Figure 7**). The LDSC confounding estimates were higher using the population GWAS data (median 44%: Q1-Q3 35%, 56%) than both the within-sibship model and previous studies. For example, the population model LDSC ratio estimates were higher for height (25%, [22%, 28%]), BMI (23%, [20%, 26%]) and educational attainment (47%, [43%, 51%]) (**Supplementary Table 11)**.

The observed non-zero confounding in the within-sibship model was unexpected because of the intuition that the within-sibship GWAS models are unlikely to be confounded. The LDSC ratios in the population GWAS were also higher than previous studies. We followed up these findings by evaluating the effects of LD score mismatch and cryptic relatedness on the LDSC ratios.

Evaluation of LD score mismatch

A large proportion of samples in the meta-analysis were from UK based studies such as UK Biobank and Generation Scotland, for which the LD scores, generated using 1000 Genomes project (phase 3) European samples (CEU, TSC, FIN, GBR), have been shown to fit reasonably well [20]. However, a large number of samples were from Scandinavian populations (HUNT, FinnTwin), where LD mismatch leading to elevated LDSC intercept/ratios has been previously discussed [20]. We investigated this possibility using empirical and simulated data.

We investigated variation in LDSC ratios across populations by comparing ratios for height across well-powered individual studies (N > 5,000): UK Biobank, HUNT, China Kadoorie Biobank (using default East Asian LD scores), Generation Scotland, DiscoverEHR, QIMR and FinnTwin. We found some evidence of heterogeneity between studies; ratio estimates were higher in Scandinavian studies compared to UK-based studies (**Supplementary Figure 8**). We also calculated within-sibship ratio estimates for BMI, SBP and educational attainment using UK Biobank summary data. UK Biobank estimates were largely consistent with zero confounding although confidence intervals were wide (**Supplementary Table 12**).

We performed simulations to evaluate potential mismatch between the Norwegian HUNT study and the default LD scores, which were generated using 1000 Genomes data. We used simulated phenotypes and real genotype data from UK Biobank and HUNT. We estimated the LDSC ratios as above, hypothesising that estimates higher than 0 are likely to reflect LD score mismatch because the phenotypes were simulated to not be influenced by confounders or common environmental terms (which could lead to cryptic relatedness).

Our process was as follows:

a) Select 1,000 HapMap3 SNPs at random.

b) Simulate beta weights for each SNP under a normal distribution with variance defined as a function of allele frequencies. The beta weight for SNP j was simulated as follows:

$Beta_j \sim N(0, 2p_j\left(1-p_j\right))$ where p_j is the minor allele frequency of SNP j.

c) Generate polygenic scores for each individual using these weights.

d) Simulate phenotype with 30% of variation explained by polygenic score, with the rest of the variation random.

e) Run GWAS on the simulated phenotype.

In UK Biobank we used the Sibling GWAS pipeline on the same sample of siblings. In HUNT we used FastGWA [52] with a sparse GRM on a sample of 30,694 individuals not included in the sibling GWAS sample. The GWAS method and study sample is not particularly important in this context as there were no common environmental effects or confounders in the simulations.

f) Apply LDSC using EUR LD scores to estimate LDSC ratios.

From 10 simulations, the median LDSC ratio estimate was 0.05 (95% C.I. [-0.02, 0.12]) in the population model and 0.05 (95% C.I. [-0.07, 0.16]) in the within-sibship model in UK Biobank, consistent with minimal confounding. In contrast, the median ratio estimate in HUNT was 0.16 (95% C.I. [0.09, 0.23]) when using the default 1000 Genomes LD scores, highly suggestive of non-zero confounding. Using a HUNT-specific LD score reference panel generated using whole genome sequencing data, the median ratio estimate decreased to 0.11 (95% C.I. [0.04, 0.20]) but still suggested non-zero confounding. However, as detailed below, this did not lead to bias in SNP h^2 estimates.

The combined findings from the empirical and simulated analyses suggest that LD score mismatch with the 1000 Genomes LD scores in HUNT and other studies likely contributed to inflated LDSC ratios in both population and within-sibship GWAS models.

Cryptic relatedness

One source of inflation in GWAS associations is cryptic relatedness; non-independence between close relatives in the study sample results which leads to inflated precision. In sibling GWAS models we clustered standard errors over sibships, but this clustering does not account for non-independence between related sibships, e.g. uncle/mother and two offspring. Inflated signal relating to cryptic relatedness may result in confounded signal, which is detected by the LD score intercept/ratio. In conventional population GWAS, close relatives are either removed or a mixed model is used to account for close relatives.

The HUNT study population includes many second- and third-degree relatives. To investigate the extent to which cryptic relatedness may have impacted LDSC ratio estimates from the population model, we investigated the effect on the LDSC ratio of using a method that accounts for relatedness. We ran a conventional population GWAS of height using FastGWA [52], which accounts for close relatives using a sparse GRM (IBD > 0.05). We included age, sex, batch and the first 20 principal components as covariates. Using the GWAS summary data we then estimated the LDSC ratio using the 1000 Genomes reference panel and compared with previously described ratio estimates. We found that the FastGWA LDSC ratio (0.33; 95% C.I. [0.28, 0.39]) was substantially lower than the population model LDSC ratio (0.69; 95% C.I. [0.65, 0.73]) suggesting that cryptic relatedness was a source of inflation in the LDSC ratio for the population model.

Cryptic relatedness is an issue for non-family models but may not be an issue for within-family models. We performed simulations to investigate how cryptic relatedness would affect the standard errors of the population and within-sibship GWAS models.

Simulations included 3 generations (generations 1, 2 and 3), and we considered only a single genetic variant *G*. We assumed random mating across all generations and complete Mendelian inheritance for G. Individuals in generation 1 were all unrelated and after pairing randomly, each pair had 2 offspring (generation 2). Similarly, individuals in generation 2 paired randomly and had 2 offspring (generation 3). Generation 2 contained sibling pairs and Generation 3 contained first cousin quads (i.e. two pairs of siblings who are first cousins).

We simulated a common environmental term *C* for Generation 2, which was identical for the full-siblings. In Generation 3, *C* was defined as the mean of parental *C* in Generation 2. We then simulated a normally distributed phenotype *P* in Generation 3 in which 30% of the variation was explained by *C* and the other 70% of variation was random. Note that *P* is not associated with *G*. We then performed regressions of the genetic variant on the phenotype using the population and within-sibship models, extracting the regression P-values. We repeated these simulations and regressions 10,000 times. We found that the type 1 error rate was inflated in the population model (5.84 %) (i.e. the false positive rate was higher than 5%) but not in the within-sibship model (4.94 %).

These findings suggest that the standard errors in the within-sibship model are not underestimated because of cryptic relatedness relating to common environmental effects shared between relatives. This, cryptic relatedness likely inflated LDSC ratios in the population models but not in the within-sibling data. Code for simulations on cryptic relatedness is available on GitHub (github.com/LaurenceHowe/SiblingGWASPost/blob/master/LDSCsimulations/CrypticRelatednessSims.R).

Within-sibship SNP heritability estimates

LDSC was used to generate SNP heritability estimates for 25 phenotypes using the LDSC harmonised (see above) meta-analysis summary data. The summary data were harmonised using the LDSC munge_sumstats.py function, and we used the precomputed European LD scores from 1000 Genomes Phase 3.

LDSC requires a sample size parameter N to estimate SNP heritability. For this parameter, we used the effective sample size for each meta-analysis phenotype, equivalent to the number of independent observations. This was estimated as follows using GWAS standard errors, minor allele frequencies and the phenotype standard deviations (after adjusting for covariates).

$$Effective\ N\ = \frac{1}{SE^2}\frac{SD\_Resid^2}{2\times MAF\times(1-MAF)}$$

SE = GWAS model standard error, MAF = minor allele frequency of the variant, SD_Resid = standard deviation of the regression residual.

Effective sample size was estimated for each individual study GWAS and each model (e.g. UK Biobank population GWAS of height). To reduce noise from low frequency variants, we restricted to variants with MAF between 0.1 and 0.4 (from 1000 Genomes EUR). At the meta-analysis stage, the effective sample size for each variant was calculated as the sum of sample sizes of studies that the variant was present in.

We used simulated data to validate the use of effective sample sizes and to explore the effects of bias in the LDSC intercept (relating to LD score mismatch) on SNP heritability estimates. In the previously described simulations (in "evaluation of LD score mismatch") we also estimated SNP heritability alongside the LDSC ratios. In UK Biobank, the median SNP heritability across 10 simulations was 0.29 (95% C.I. [0.23, 0.34]) in the population model and 0.32 (95% C.I. [0.21, 0.42]) in the within-sibship model, highly comparable to the true simulated heritability of 0.30. In HUNT, SNP heritabilities were unbiased using both reference panels, but the median SNP heritability estimate was more precise using the HUNT LD scores (0.31; 95% C.I. [0.25, 0.38]) than the 1000 Genomes LD scores (0.31; 95% C.I. [0.21, 0.42]).

The simulated data suggests that LDSC can generate unbiased estimates of SNP heritability even in the presence of LD score mismatch. However, extensive simulations beyond the scope of this project are required to investigate this further. In empirical analyses, we decided to focus on the differences between the population model (h^2_{Pop}) and within-sibship model (h^2_{WS}) SNP heritability estimates. If we assume that biases affect the estimates equally then the difference between the two estimates will be unbiased. We estimated the difference between the heritability estimates (h^2_{Diff}) using a difference-of-two-means test [53] as below.

$$h^2_{Diff} = h^2_{Pop} - h^2_{WS}$$

$$SE\left(h^2_{Diff}\right) \sim \sqrt{\mathrm{SE}(h^2_{Pop})^2 + \mathrm{SE}(h^2_{WS})^2 - 2Cov(h^2_{Pop}, h^2_{WS})}$$

To estimate $Cov(h^2_{Pop}, h^2_{WS})$, we computed the cross-GWAS LDSC intercept between the population and within-sibship GWAS data (for the same phenotype) which is an estimate of $Cor(h^2_{Pop}, h^2_{WS})$. The estimates of this term were ~0.40 across phenotypes. We then calculated the covariance term as follows:

$$Cov(h^2_{Pop}, h^2_{WS}) = Cor(h^2_{Pop}, h^2_{WS})\ \times\ \mathrm{SE}(h^2_{Pop}) \times \mathrm{SE}(h^2_{WS})$$

We used the difference Z score (i.e $\frac{h^2_{Diff}}{SE\left(h^2_{Diff}\right)}$) to generate a P-value for the difference between $h^2_{Pop}$and h^2_{WS}. In the text, we report differences reaching nominal significance (difference P < 0.05).

We calculated the expected effect of shrinkage on LDSC SNP heritability estimates. LDSC heritability estimates (h^2) are derived from the formulation below [20]:

$$\chi^2 \sim \frac{Nh^2 l_j}{M} + Na + 1$$

where χ^2 = the square of the GWAS Z score, N = the sample size, M = number of variants such that $\frac{h^2}{M}$ is the average heritability for each variant, l_j is the LD score of variant j, a is the effect of confounding biases.

Uniform shrinkage across the genome, would lead to GWAS Z scores being multiplied by a factor $(1-k)$ where k is the shrinkage coefficient and χ^2 statistics being multiplied by $(1-k)^2$. As above, we have used effective sample size to account for differences in N between the population and within-sibship models. Therefore, assuming all other coefficients remain consistent,the expectation of h^2_{WS} can be written as a function of k and h^2_{Pop}.

$$h^2_{Pop} = y$$

$$h^2_{WS} = (1-k)^2 y$$

Within-sibship genetic correlations with educational attainment

We used LDSC to estimate r_g between educational attainment and other phenotypes using both population and within-sibship data. LDSC requires non-zero heritability to generate meaningful r_g estimates, so we restricted analyses to the 22 phenotypes with SNP heritability point estimates greater than zero in both population and within-sibship models. We estimated the difference between the population ($r_{g,Pop}$) and within-sibship ($r_{g,WS}$) estimates ($r_{g,Diff}$) using a difference-of-two-means test [53].

$$r_{g,Diff} = r_{g,Pop} - r_{g,WS}$$

We used the jackknife to estimate the standard error of the difference $SE(r_{g,Diff})$. After restricting to ~1.2 million Hapmap 3 variants present in the 1000 Genomes LD scores, we ordered variants by chromosome and base-pair and separated variants into 100 blocks. We removed each block in turn and computed $r_{g,Diff}$ using LDSC 100 times. We then calculated $SE(r_{g,Diff})$ across the 100 iterations as follows:

$$SE\left(r_{g,Diff}\right) = \sqrt{\frac{99}{100}\sum_{1}^{100}\left(r_{g,Diff_k} - \mu\right)^2}$$

where $\mu = \frac{\sum_1^{100} r_{g,Diff,k}}{100}$

$r_{g,Diff,k}$= r_g estimate in the kth iteration, μ = the mean r_g estimate across all 100 iterations

We used the difference Z score (i.e. $\frac{r_{g,Diff}}{SE(r_{g,Diff})}$) to generate a P-value for heterogeneity between $r_{g,Pop}$ $r_{g,WS}$. In the text, we report differences reaching nominal significance (heterogeneity P < 0.05).

Within-sibship Mendelian randomization: effects of height and BMI

We performed Mendelian randomization analyses using the within-sibship meta-analysis GWAS data to estimate the effect of two exposures (height and BMI) on 23 outcome phenotypes. For the exposure instruments, we used 803 and 418 independent genetic variants for height and BMI, respectively. These variants were identified by LD clumping in PLINK (r^2 < 0.001, physical distance threshold = 10,000 kb, P < $5x10^{-8}$) as described in the PRS analysis. We then performed a MR-IVW analysis using the within-sibship meta-analysis data to estimate the effect of the exposure on the outcome as

$$\beta_{MR} = \sum \frac{\beta_{Exp} * \beta_{Out}}{(\sigma_{Out})^2} / \sum \frac{(\beta_{Exp})^2}{(\sigma_{Out})^2}$$

$where\ \beta_{Exp}$ = association estimate from exposure GWAS, β_{Out} = association estimate from outcome GWAS, σ_{Out} = standard error from outcome GWAS.

We also performed Mendelian randomization analyses using the population meta-analysis GWAS data for comparison. We estimated differences between population and within-sibship Mendelian randomization estimates using the difference-of-two-means test [53]:

$$\beta_{MR,Diff} = \beta_{MR,Pop} - \beta_{MR,WS}$$

We used the jackknife to estimate the standard error of the difference $SE(\beta_{MR,Diff})$. With n genetic instruments, we removed each variant from the analysis in turn and then computed $\beta_{MR,Diff}$, storing the estimate from the n iterations. We then calculated $SE(\beta_{MR,Diff})$ as follows:

$$SE(\beta_{MR,Diff}) = \sqrt{\frac{n-1}{n}\sum_{1}^{n}(\beta_{MR,Diff,k} - \mu)^2}$$

$$\text{where } \mu = \frac{\sum_1^n \beta_{MR,Diff,k}}{n}$$

n = number of genetic variants used as instruments, $\beta_{MR,Diff,k}$= $\beta_{MR,Diff}$ estimate in the kth iteration, μ = the mean $\beta_{MR,Diff}$ estimate across all n iterations.

We used the difference Z score (i.e. $\frac{\beta_{MR,Diff}}{SE(\beta_{MR,Diff})}$) to generate a P-value for heterogeneity between $\beta_{MR,Pop}$ and $\beta_{MR,WS}$. In the text, we report differences reaching nominal significance (heterogeneity P < 0.05).

Polygenic adaptation

Polygenic adaptation was estimated using similar methods to a previous publication [28]. Precomputed SDS scores were downloaded for UK10K data from https://web.stanford.edu/group/pritchardlab/ Genomic regions under strong recent selection (*MHC* chr6: 25,892,529-33,436,144; lactase chr2: 134,608,646-138,608,646) were removed and SDS scores were normalised within each 1% allele frequency bin.

SDS scores were merged with GWAS meta-analysis data for 25 phenotypes. Variants with low effective sample sizes (< 50% of maximum) were removed for each phenotype. SDS scores were transformed to tSDS such that the reference allele was the phenotype-increasing allele.

Spearman's rank test was used to estimate the correlation between tSDS and the absolute value of GWAS Z-scores from the population and within-sibship models. Standard errors were estimated using the jackknife. The genome was ordered by chromosome and base pair and divided into 100 blocks. Correlations were estimated 100 times with each kth block removed in turn. The standard error of the correlation estimate $SE(Cor)$ was calculated as follows:

$$SE(Cor) = \sqrt{\frac{99}{100}\sum_{1}^{100}(Cor_k - \mu)^2}$$

$$\text{where } \mu = \frac{\sum_1^{100} Cor_k}{100}$$

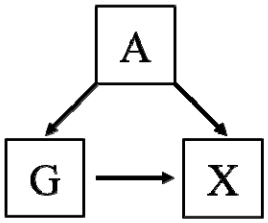
Cor_k= Spearman's rank correlation estimate in the kth iteration, μ = the mean correlation estimate across the 100 iterations.

Given previous concerns [26, 27], we performed several sensitivity analyses for the height analysis. First, we evaluated the mean tSDS in the within-sibship model for a subset of independent variants strongly associated with height. We determined these variants by LD clumping the within-sibship meta-analysis GWAS data in PLINK v 2.0 ($r^2 < 0.001$, physical distance threshold = 10,000 kb, $P < 1\times10^{-5}$). Second, we used LDSC to estimate the genetic correlation between the SDS scores and the height GWAS data from the population and within-sibship models. The SDS input data was normalised (as above) SDS and we used the precomputed European LD scores from 1000 Genomes. Third, we also calculated spearman rank correlations between height and the SDS (as above) using summary data from individual studies (as opposed to the meta-analysis GWAS). We used all studies with N > 4000, which were UK Biobank, HUNT, Generation Scotland, QIMR, Netherlands Twin Registry, FinnTwin, Discover EHR and China Kadoorie Biobank and investigated both population/within-sibship models. We then used a fixed effects model to meta-analyse the correlation estimates across the studies for the population and within-sibship models. Notably, the correlation estimate using only the UK Biobank WF summary data was inconclusive (r = 0.002; 95% C.I. -0.005, 0.010), consistent with a previous study [26], and correlation point estimates from individual studies were generally smaller than the meta-analysis GWAS estimates. This heterogeneity could relate to the increased number of samples in the meta-analysis, with a higher signal to noise ratio in the individual studies.

Figures

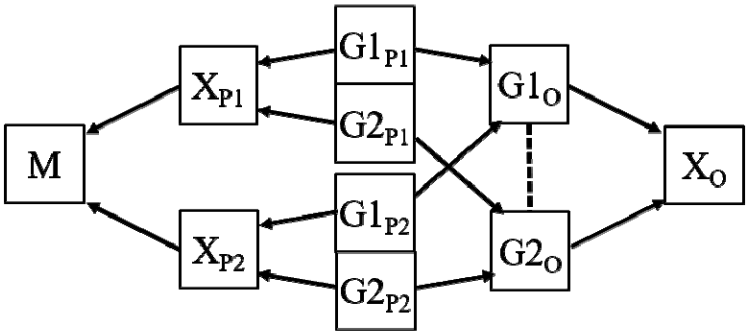
Figure 1A Demographic and indirect genetic effects

Fine-scale population structure



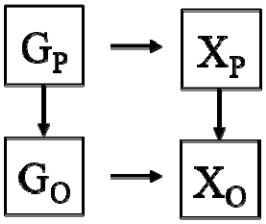
A = Ancestry, G = Genotype, X = Phenotype.

Assortative mating



$G1_{P1, P2, O}$ = Genotypes of Parent1, Parent2 and Offspring for variant 1,
$G2_{P1, P2, O}$ = Genotypes for variant 2,
$X_{P1, P2, O}$ = Phenotypes,
M = Mate choice.

Indirect genetic effects (parental)



G_P = Parental genotype, G_O = Offspring genotype,
X_P = Parental phenotype, X_O = Offspring phenotype.

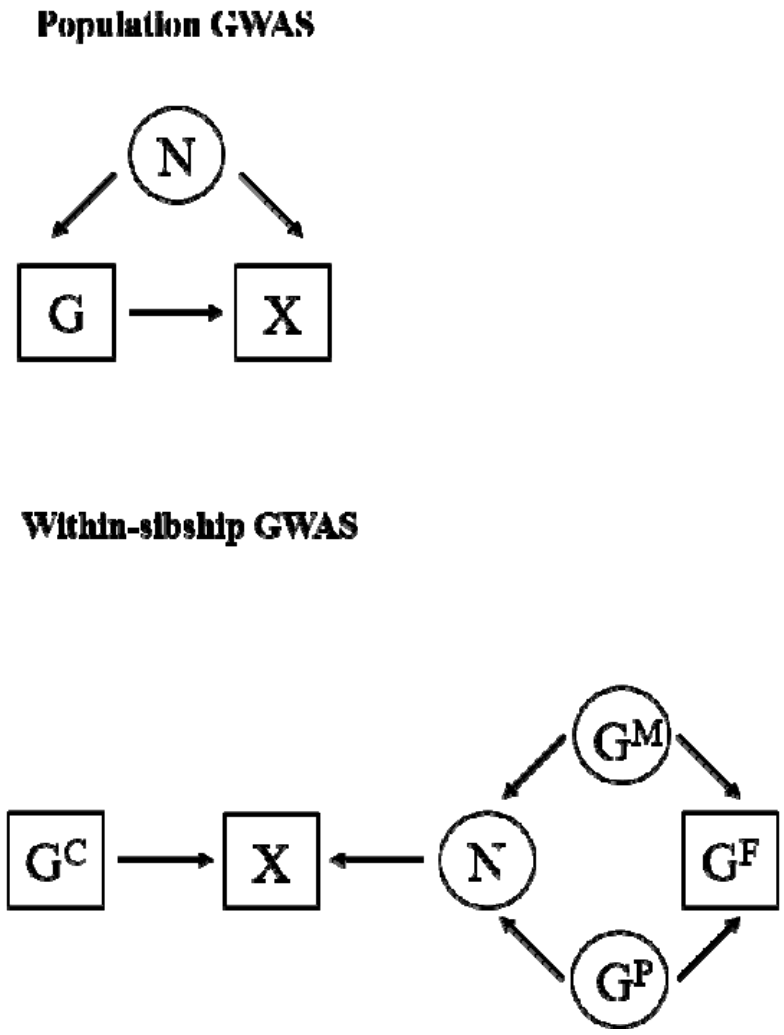
Description for Figure 1A

Population stratification bias: Population stratification is defined as the distortion of associations between a genotype and a phenotype when ancestry *A* influences both genotype *G* (via differences in allele frequencies) and the phenotype *X*. Principal components and linear mixed model methods control for ancestry but there may be residual confounding relating to fine-scale population structure.

Assortative mating: Assortative mating is a phenomenon where individuals select a partner based on phenotypic (dis)similarities. For example, tall individuals may prefer a tall partner. Assortative mating can induce correlations between causes of an assorted phenotype in subsequent generations. If a phenotype *X* is influenced by 2 independent genetic variants *G1* and *G2* then assortment on *X* (represented by effects of *X* on mate choice *M*) will induce positive correlations between *G1* in parent 1 and *G2* in parent 2 and vice versa. Parental transmission will then induce correlations between otherwise independent *G1* and *G2* in offspring. These correlations can distort genetic association estimates.

Indirect genetic effects: Indirect genetic effects are effects of relative genotypes (via relative phenotypes and the shared environment) on the index individual's phenotype. These indirect effects influence population GWAS estimates because relative genotypes are also associated with genotypes of the index individual. Indirect genetic effects of parents on offspring are of most interest because they are likely to be the largest. However, indirect genetic effects of siblings or more distal relatives are also possible.

Figure 1B Population and within-sibship GWAS



Description for Figure 1B

Population GWAS estimate the association between raw genotypes *G* and phenotypes *X*. As outlined in Figure 1A, estimates from population GWAS may not fully control for effects of demographic biases (population stratification and assortative mating) and may also capture indirect genetic effects of relatives. For simplicity we use *N* to represent all sources of associations between *G* and *X* which do not relate to direct effects of *G*. Circles indicate unmeasured variables and squares indicate measured variables.

If parental genotypes are known, *G* can be separated into non-random (determined by parental genotypes) and random (relating to segregation at meiosis) components.

Within-sibship GWAS include the mean genotype across a sibship (G^F) (a proxy for the mean of the paternal and maternal genotypes $G^{P, M}$) as a covariate to capture associations between *G* and *X* relating to parents. The within-sibship estimate is defined as the effect of the random component; i.e. the association between family-mean centred genotype G^C (i.e. $G - G^F$) and *X*. Demographic

Figure 1C Flowchart illustrating datasets and analyses

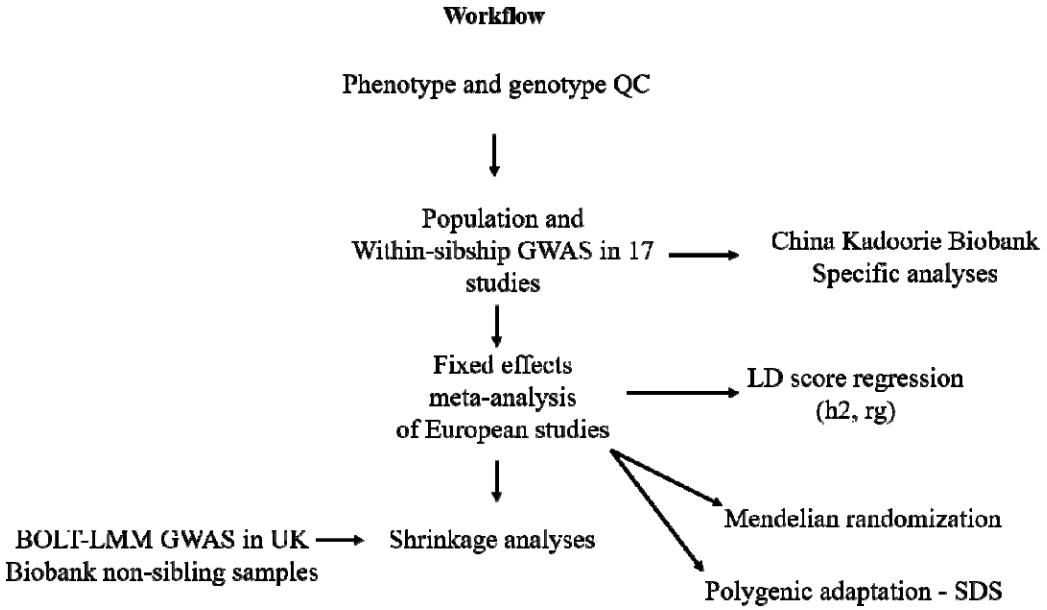
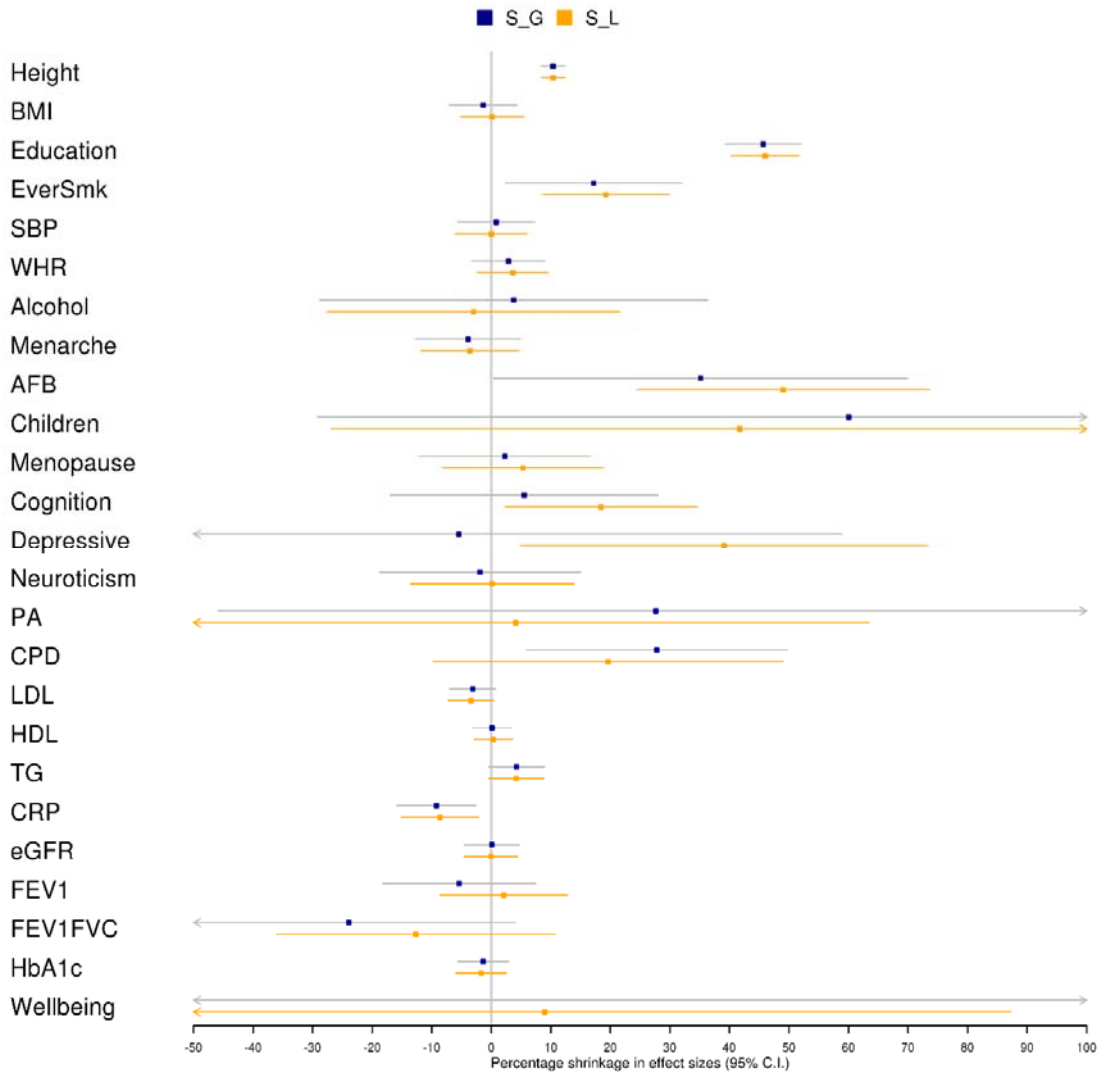


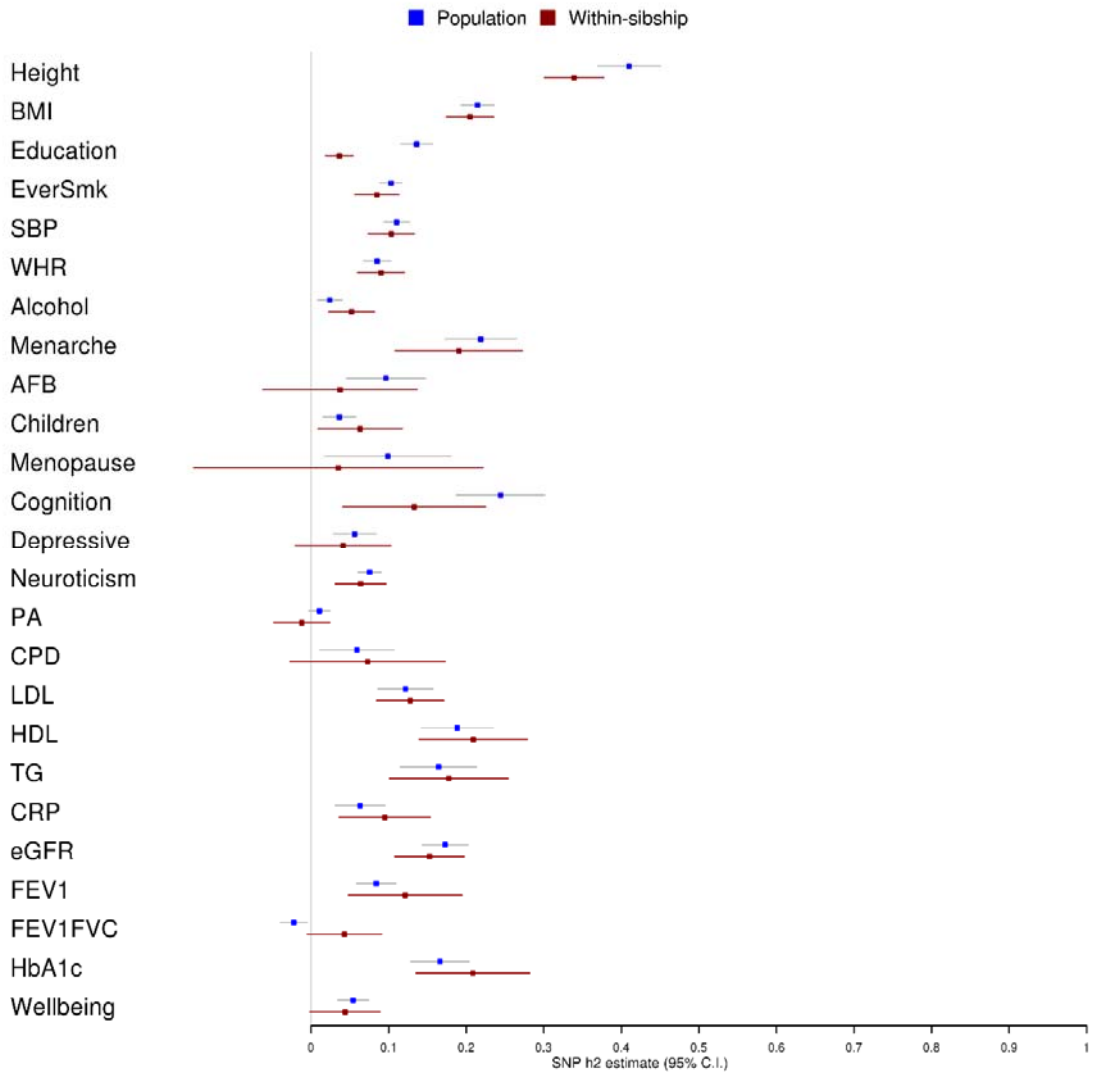
Figure 2 Effect size shrinkage in within-sibship models



S_G = weighted score at genome-wide significance ($P < 5x10^{-8}$), S_L, = weighted score at more liberal threshold ($P < 1x10^{-5}$) BMI = body mass index, Education = educational attainment, EverSmk = ever smoking, SBP = systolic blood pressure, WHR = waist-hip ratio, Alcohol = weekly alcohol consumption, Menarche = age at menarche, AFB = age at first birth, Children = number of biological children, Menopause = age at menopause, Cognition = cognitive ability, Depressive = depressive symptoms, PA = physical activity, CPD = cigarettes per day, LDL = LDL cholesterol, HDL = HDL cholesterol, TG = triglycerides, CRP = C-reactive protein, eGFR = estimated glomerular filtration rate, FEV1 = forced expiratory volume, FEV1FVC = ratio of FEV1/forced vital capacity, HbA1c = Haemoglobin A1C.

Figure 2 displays estimates of shrinkage between population and within-sibship models. Shrinkage is defined as the % decrease in association between the relevant weighted score and phenotype when comparing the population estimate to the within-sibship estimate.

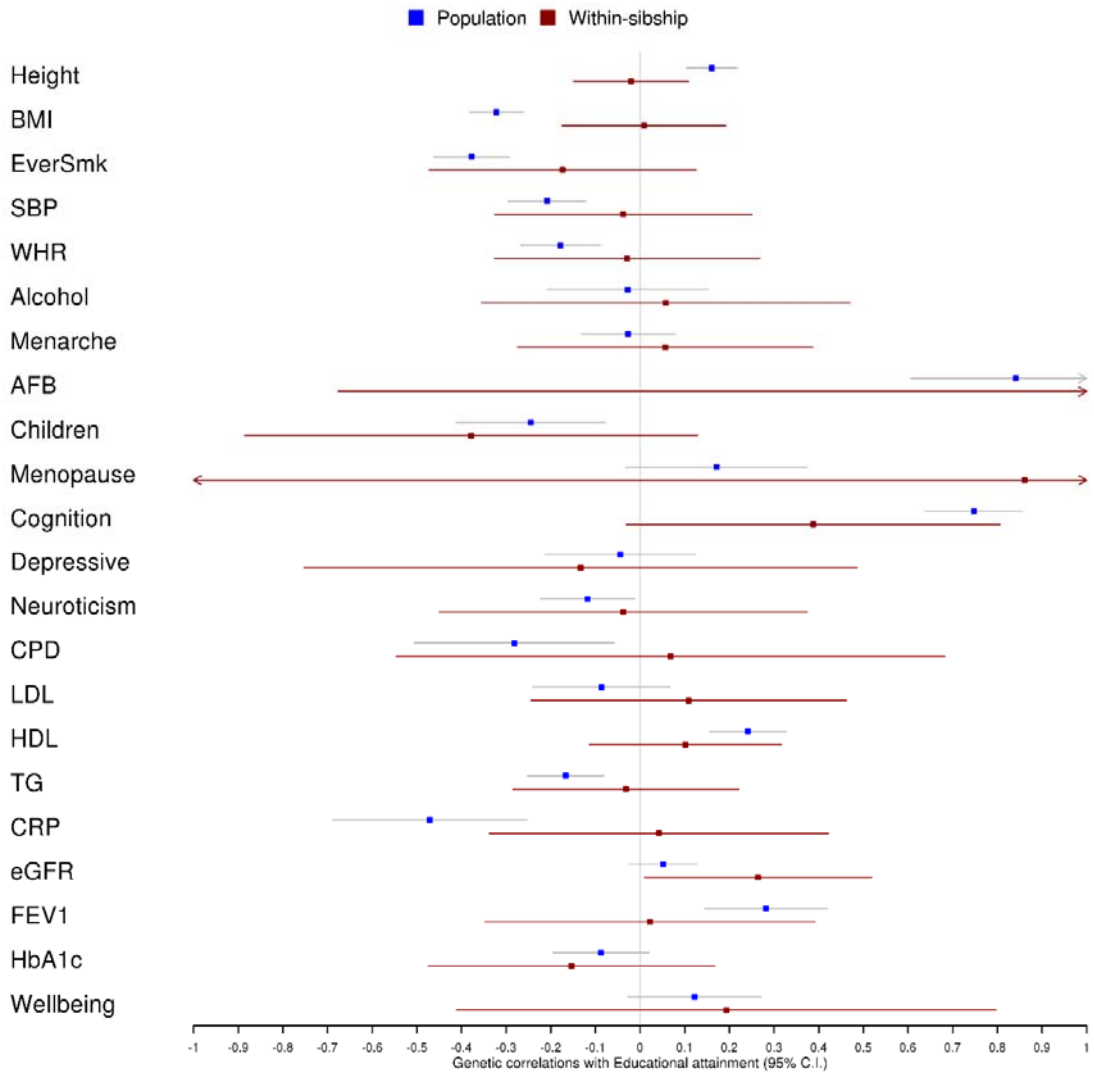
Figure 3 Observed-scale LDSC SNP heritability estimates for 25 phenotypes



BMI = body mass index, Education = educational attainment, EverSmk = ever smoking, SBP = systolic blood pressure, WHR = waist-hip ratio, Alcohol = weekly alcohol consumption, Menarche = age at menarche, AFB = age at first birth, Children = number of biological children, Menopause = age at menopause, Cognition = cognitive ability, Depressive = depressive symptoms, PA = physical activity, CPD = cigarettes per day, LDL = LDL cholesterol, HDL = HDL cholesterol, TG = triglycerides, CRP = C-reactive protein, eGFR = estimated glomerular filtration rate, FEV1 = forced expiratory volume, FEV1FVC = ratio of FEV1/forced vital capacity, HbA1c = Haemoglobin A1C.

Figure 3 displays LDSC SNP h^2 estimates for 25 phenotypes using population and within-sibship meta-analysis data.

Figure 4 LDSC genetic correlations of phenotypes with educational attainment



BMI = body mass index, Education = educational attainment, EverSmk = ever smoking, SBP = systolic blood pressure, WHR = waist-hip ratio, Alcohol = weekly alcohol consumption, Menarche = age at menarche, AFB = age at first birth, Children = number of biological children, Menopause = age at menopause, Cognition = cognitive ability, Depressive = depressive symptoms, CPD = cigarettes per day, LDL = LDL cholesterol, HDL = HDL cholesterol, TG = triglycerides, CRP = C-reactive protein, eGFR = estimated glomerular filtration rate, FEV1 = forced expiratory volume, HbA1c = Haemoglobin A1C.

Figure 4 displays LDSC rg estimates between educational attainment and 22 phenotypes using population and within-sibship meta-analysis data.

Figure 5 Individual study, pooled and GWAS meta-analysis estimates for polygenic adaptation on height

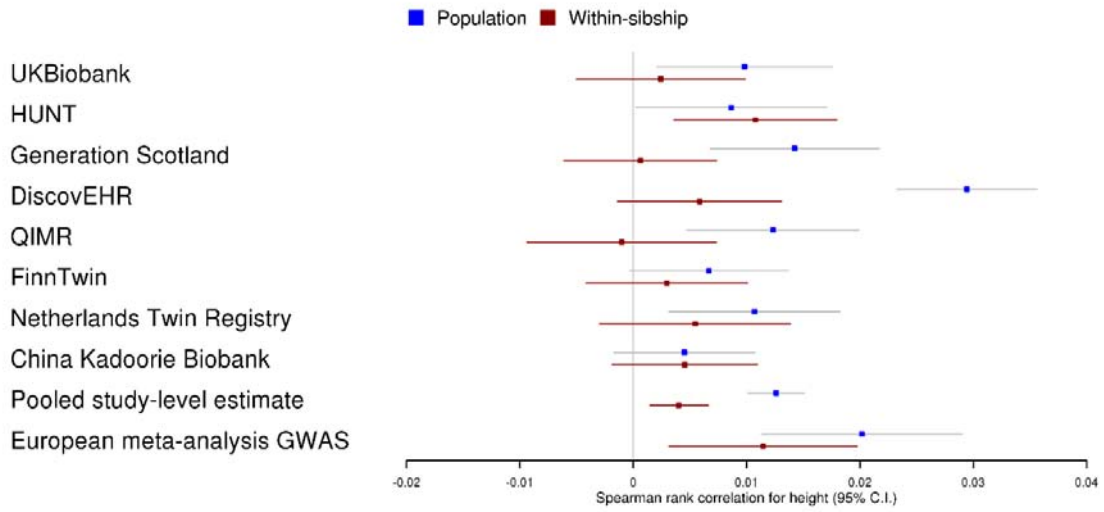


Figure 5 displays spearman rank correlation estimates between tSDS (SDS scores aligned with height increasing alleles) and absolute height Z scores. Positive correlations indicate evidence of historical positive selection on height increasing alleles. The pooled estimate is a meta-analysis of the correlation estimates from the individual studies shown above while the European meta-analysis estimate is the correlation estimate using the meta-analysis GWAS data.

Tables

Table 1 Within-sibship Mendelian randomization: effects of height and BMI on 23 phenotypes

Outcome (units)	**IVW estimate of effect of SD increase in height on outcome (95% C.I.)**		**Diff P**	**IVW estimate of effect of SD increase in BMI on outcome (95% C.I.)**		**Diff P**
	Population	Within-sibship		Population	Within-sibship	
Age at first birth (years)	0.24 (0.09, 0.38)	0.08 (-0.17, 0.32)	0.16	-0.88 (-1.18, -0.58)	-0.20 (-0.64, 0.23)	< 0.001
Alcohol consumption (units)	0.02 (-0.04, 0.08)	0.02 (-0.08, 0.13)	0.89	-0.14 (-0.26, -0.01)	-0.20 (-0.39, -0.01)	0.45
Cigarettes per day	0.27 (0.04, 0.50)	0.40 (-0.02, 0.81)	0.52	0.68 (0.18, 1.17)	0.46 (-0.29, 1.21)	0.56
C-reactive protein (SD)	-0.03 (-0.05, -0.01)	-0.01 (-0.04, 0.03)	0.12	0.28 (0.24, 0.32)	0.23 (0.17, 0.30)	0.13
Number of children	-0.01 (-0.04, 0.01)	0.01 (-0.04, 0.06)	0.29	0.08 (0.02, 0.14)	0.05 (-0.04, 0.14)	0.52
Cognitive ability (SD)	0.07 (0.03, 0.10)	0.05 (-0.00, 0.10)	0.46	-0.19 (-0.26, -0.12)	-0.11 (-0.20, -0.02)	0.12
Depressive symptoms (SD)	-0.01 (-0.04, 0.01)	-0.02 (-0.07, 0.02)	0.63	0.03 (-0.03, 0.08)	-0.01 (-0.09, 0.07)	0.39
Educational attainment (SD)	0.05 (0.03, 0.06)	0.01 (-0.01, 0.03)	0.0029	-0.18 (-0.21, -0.14)	-0.04 (-0.08, 0.00)	< 0.001
Ever smoking (risk difference)	-0.01 (-0.01, 0.00)	0.00 (-0.01, 0.02)	0.08	0.06 (0.05, 0.08)	0.04 (0.02, 0.07)	0.12
FEV1 (SD)	-0.03 (-0.05, -	-0.04 (-0.08, 0.00)	0.59	-0.17 (-0.22, -	-0.17 (-0.25, -	0.92

	0.01)			0.12)	0.10)	
FEV1FVC (SD)	0.02 (-0.00, 0.04)	0.02 (-0.02, 0.06)	0.90	-0.02 (-0.06, 0.02)	-0.02 (-0.09, 0.05)	0.97
HbA1c (SD)	-0.00 (-0.02, 0.02)	0.02 (-0.02, 0.06)	0.18	0.14 (0.09, 0.18)	0.14 (0.07, 0.21)	0.91
HDL cholesterol (SD)	-0.01 (-0.03, 0.01)	-0.02 (-0.05, 0.00)	0.24	-0.32 (-0.35, -0.28)	-0.31 (-0.36, -0.27)	0.91
LDL cholesterol (SD)	-0.05 (-0.06, -0.03)	-0.03 (-0.06, 0.00)	0.25	0.03 (-0.01, 0.07)	0.04 (-0.01, 0.09)	0.68
Age at menarche (Years)	0.08 (0.04, 0.12)	0.07 (-0.00, 0.14)	0.72	-0.61 (-0.70, -0.52)	-0.60 (-0.72, -0.47)	0.87
Age at menopause (Years)	-0.17 (-0.36, 0.03)	-0.13 (-0.48, 0.22)	0.81	-0.48 (-0.90, -0.07)	-0.24 (-0.87, 0.38)	0.50
Neuroticism (SD)	-0.01 (-0.03, 0.00)	0.01 (-0.02, 0.04)	0.09	0.01 (-0.02, 0.05)	-0.04 (-0.09, 0.02)	0.11
Physical activity (risk difference)	-0.00 (-0.01, 0.01)	-0.01 (-0.03, 0.01)	0.17	-0.03 (-0.05, -0.01)	-0.02 (-0.05, 0.01)	0.55
SBP (mmHg)	-0.81 (-1.08, -0.54)	-0.69 (-1.16, -0.22)	0.59	2.96 (2.37, 3.54)	2.90 (2.05, 3.74)	0.88
Triglycerides (SD)	-0.02 (-0.03, -0.00)	0.01 (-0.02, 0.04)	0.04	0.27 (0.23, 0.30)	0.26 (0.21, 0.32)	0.88
Waist-hip ratio adj for BMI (WHR*100)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.18	0.01 (0.01, 0.01)	0.01 (0.01, 0.01)	0.52
Wellbeing (SD)	0.00 (-0.02, 0.02)	-0.02 (-0.05, 0.02)	0.28	-0.05 (-0.10, -0.01)	-0.05 (-0.11, 0.02)	0.84
eGFR	-0.69 (-0.93, -	-0.85 (-1.26, -	0.43	-0.06 (-0.57, 0.46)	0.15 (-0.59, 0.88)	0.56

	0.46)	0.45)				

SBP = systolic blood pressure, eGFR = estimated glomerular filtration rate.

Table 1 contains population and within-sibship Mendelian randomization estimates of height and BMI on 23 phenotypes. Units are presented in terms of a standard deviation increase in height or BMI. Difference (diff) P-values refer to evidence of differences between population and within-sibship estimates.

References

1. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. The American Journal of Human Genetics. 2017;101(1):5-22.
2. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Communications biology. 2019;2(1):9. doi: 10.1038/s42003-018-0261-x.
3. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273(5281):1516-7.
4. Morris TT, Davies NM, Hemani G, Davey Smith G. Population phenomena inflate genetic associations of complex social traits. Science advances. 2020;6(16):eaay0328. Epub 2020/05/20. doi: 10.1126/sciadv.aay0328. PubMed PMID: 32426451; PubMed Central PMCID: PMCPMC7159920.
5. Fisher RA. The Genetical Theory of Natural Selection: Oxford Univ. Press; 1930.
6. Young AI, Nehzati SM, Lee C, Benonisdottir S, Cesarini D, Benjamin DJ, et al. Mendelian imputation of parental genotypes for genome-wide estimation of direct and indirect genetic effects. biorXiv. 2020:2020.07.02.185199. doi: 10.1101/2020.07.02.185199 %J bioRxiv.
7. Brumpton B, Sanderson E, Hartwig FP, Harrison S, Vie GÅ, Cho Y, et al. Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. Nature Communications. 2020:602516.
8. Shen H, Feldman MW. Genetic nurturing, missing heritability, and causal analysis in genetic statistics. Proceedings of the National Academy of Sciences. 2020;117(41):25646-54. doi: 10.1073/pnas.2015869117.
9. Howe LJ, Lawson DJ, Davies NM, Pourcain BS, Lewis SJ, Davey Smith G, et al. Genetic evidence for assortative mating on alcohol consumption in the UK Biobank. Nature Communications. 2019. doi: 10.1038/s41467-019-12424-x.
10. Robinson MR, Kleinman A, Graff M, Vinkhuyzen AA, Couper D, Miller MB, et al. Genetic evidence of assortative mating in humans. Nature Human Behaviour. 2017;1:0016.
11. Yengo L, Robinson MR, Keller MC, Kemper KE, Yang Y, Trzaskowski M, et al. Imprint of assortative mating on the human genome. Nature Human Behaviour. 2018:300020.
12. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. Nature Communications. 2019;10(1):333.
13. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, et al. The nature of nurture: Effects of parental genotypes. Science. 2018;359(6374):424-8. doi: 10.1126/science.aan6877 %J Science.
14. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50(8):1112-21. Epub 2018/07/25. doi: 10.1038/s41588-018-0147-3. PubMed PMID: 30038396; PubMed Central PMCID: PMCPMC6393768.
15. Warrington NM, Freathy RM, Neale MC, Evans DM. Using structural equation modelling to jointly estimate maternal and fetal effects on birthweight in the UK Biobank. Int J Epidemiol. 2018;47(4):1229-41. Epub 2018/02/16. doi: 10.1093/ije/dyy015. PubMed PMID: 29447406; PubMed Central PMCID: PMCPMC6124616.
16. Warrington NM, Beaumont RN, Horikoshi M, Day FR, Helgeland Ø, Laurin C, et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. Nat Genet. 2019;51(5):804-14. Epub 2019/05/03. doi: 10.1038/s41588-019-0403-1. PubMed PMID: 31043758; PubMed Central PMCID: PMCPMC6522365.
17. Young AI, Benonisdottir S, Przeworski M, Kong A. Deconstructing the sources of genotype-phenotype associations in humans. Science. 2019;365(6460):1396-400. doi: 10.1126/science.aax3710 %J Science.
18. Balbona J, Kim Y, Keller MC. Estimation of parental effects using polygenic scores. bioRxiv. 2020:2020.08.11.247049. doi: 10.1101/2020.08.11.247049 %J bioRxiv.
19. Selzam S, Ritchie SJ, Pingault JB, Reynolds CA, O'Reilly PF, Plomin R. Comparing Within- and Between-Family Polygenic Score Prediction. Am J Hum Genet. 2019;105(2):351-63. Epub 2019/07/16. doi: 10.1016/j.ajhg.2019.06.006. PubMed PMID: 31303263; PubMed Central PMCID: PMCPMC6698881.
20. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics. 2015;47(3):291-5.

21. Speed D, Balding DJ. SumHer better estimates the SNP heritability of complex traits from summary statistics. Nat Genet. 2019;51(2):277-84. Epub 2018/12/05. doi: 10.1038/s41588-018-0279-5. PubMed PMID: 30510236; PubMed Central PMCID: PMCPMC6485398.
22. Young Al, Frigge ML, Gudbjartsson DF, Thorleifsson G, Bjornsdottir G, Sulem P, et al. Relatedness disequilibrium regression estimates heritability without environmental bias. Nature Genetics. 2018;50(9):1304-10. doi: 10.1038/s41588-018-0178-9.
23. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nature Genetics. 2015;47(11):1236-41.
24. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology. 2003;32(1):1-22.
25. Davies NM, Howe LJ, Brumpton B, Havdahl A, Evans DM, Davey Smith G. Within family Mendelian randomization studies. Hum Mol Genet. 2019;28(R2):R170-r9. Epub 2019/10/28. doi: 10.1093/hmg/ddz204. PubMed PMID: 31647093.
26. Berg JJ, Harpak A, Sinnott-Armstrong N, Joergensen AM, Mostafavi H, Field Y, et al. Reduced signal for polygenic adaptation of height in UK Biobank. eLife. 2019;8. Epub 2019/03/22. doi: 10.7554/eLife.39725. PubMed PMID: 30895923.
27. Sohail M, Maier RM, Ganna A, Bloemendal A, Martin AR, Turchin MC, et al. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. eLife. 2019;8. Epub 2019/03/22. doi: 10.7554/eLife.39702. PubMed PMID: 30895926.
28. Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, Golan D, et al. Detection of human adaptation during the past 2000 years. 2016;354(6313):760-4. doi: 10.1126/science.aag0776 %J Science.
29. Fulker DW, Cherny SS, Sham PC, Hewitt JK. Combined linkage and association sib-pair analysis for quantitative traits. Am J Hum Genet. 1999;64(1):259-67. Epub 1999/01/23. doi: 10.1086/302193. PubMed PMID: 9915965; PubMed Central PMCID: PMCPMC1377724.
30. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. Am J Hum Genet. 2000;66(1):279-92. Epub 2000/01/13. doi: 10.1086/302698. PubMed PMID: 10631157; PubMed Central PMCID: PMCPMC1288332.
31. Pingault J-B, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using genetic data to strengthen causal inference in observational research. Nature Reviews Genetics. 2018;19(9):566-80. doi: 10.1038/s41576-018-0020-3.
32. Neale MC, Cherny SS, Sham PC, Whitfield JB, Heath AC, Birley AJ, et al. Distinguishing Population Stratification from Genuine Allelic Effects with Mx: Association of ADH2 with Alcohol Consumption. Behavior Genetics. 1999;29(4):233-43. doi: 10.1023/A:1021638122693.
33. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203.
34. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, et al. Cohort Profile: the HUNT Study, Norway. Int J Epidemiol. 2013;42(4):968-77. Epub 2012/08/11. doi: 10.1093/ije/dys095. PubMed PMID: 22879362.
35. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol. 2013;42(3):689-700. Epub 2012/07/13. doi: 10.1093/ije/dys084. PubMed PMID: 22786799.
36. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol. 2011;40(6):1652-66. Epub 2011/12/14. doi: 10.1093/ije/dyr120. PubMed PMID: 22158673; PubMed Central PMCID: PMCPMC3235021.
37. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. eLife. 2020;9. Epub 2020/01/31. doi: 10.7554/eLife.48376. PubMed PMID: 31999256; PubMed Central PMCID: PMCPMC7067566.
38. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine. 2008;27(8):1133-63.
39. Lawlor D, Richmond R, Warrington N, McMahon G, Smith GD, Bowden J, et al. Using Mendelian randomization to determine causal effects of maternal pregnancy (intrauterine) exposures on offspring outcomes: Sources of bias and methods for assessing them. Wellcome Open Research. 2017;2.
40. Hwang L-D, Tubbs JD, Luong J, Lundberg M, Moen G-H, Wang G, et al. Estimating indirect parental genetic effects on offspring phenotypes using virtual parental genotypes derived from sibling and half sibling pairs. PLOS Genetics. 2020;16(10):e1009154. doi: 10.1371/journal.pgen.1009154.

41. Silventoinen K, Jelenkovic A, Sund R, Latvala A, Honda C, Inui F, et al. Genetic and environmental variation in educational attainment: an individual-based analysis of 28 twin cohorts. Scientific Reports. 2020;10(1):12681. doi: 10.1038/s41598-020-69526-6.
42. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nature Reviews Genetics. 2002;3(11):872-82.
43. Maes HH, Prom-Wormley E, Eaves LJ, Rhee SH, Hewitt JK, Young S, et al. A Genetic Epidemiological Mega Analysis of Smoking Initiation in Adolescents. Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco. 2017;19(4):401-9. Epub 2016/11/04. doi: 10.1093/ntr/ntw294. PubMed PMID: 27807125; PubMed Central PMCID: PMCPMC5896552.
44. Stulp G, Simons MJ, Grasman S, Pollet TV. Assortative mating for human height: A meta-analysis. American journal of human biology : the official journal of the Human Biology Council. 2017;29(1). Epub 2016/09/18. doi: 10.1002/ajhb.22917. PubMed PMID: 27637175; PubMed Central PMCID: PMCPMC5297874.
45. Kong A, Benonisdottir S, Young AI. Family Analysis with Mendelian Imputations. biorXiv. 2020:2020.07.02.185181. doi: 10.1101/2020.07.02.185181 %J bioRxiv.
46. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
47. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nature Genetics. 2015;47(3):284-90.
48. Mitchell RE, Hemani G, Dudding T, Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, version 1, 13/11/2017 2017.
49. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 2007;81(3):559-75.
50. Palla L, Dudbridge F. A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. The American Journal of Human Genetics. 2015;97(2):250-9.
51. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic Epidemiology. 2013;37(7):658-65.
52. Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, et al. A resource-efficient tool for mixed model association analysis of large-scale data. Nature Genetics. 2019.
53. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ. 2003;326(7382):219. doi: 10.1136/bmj.326.7382.219 %J BMJ.

Acknowledgements

We would like to thank Hakhamanesh Mostafavi and Jonathan Pritchard for helpful suggestions and guidance relating to the polygenic adaptation analyses.

Funding

LJH, TTM, YC, DAL, GDS, GH and NMD work in a unit that receives support from the University of Bristol and the UK Medical Research Council (MC_UU_00011). MGN is supported by ZonMW grants 849200011 and 531003014 from The Netherlands Organisation for Health Research and Development, a VENI grant awarded by NWO (VI.Veni.191G.030) on NIH grant R01MH120219 and a Jacobs foundation research fellowship. DAL is PI of the Bristol British Heart Foundation Accelerator Award (AA/18/7/34219), is a British Heart Foundation Chair and National Institute of Health Research Senior Investigator (NF-0616-10102). JK receives support from Academy of Finland grants (#308248 & 312073). EMTD and KPH receive support from NIH grants (R01HD083613, R01HD092548). LK receives support from a RCUK Innovation Fellowship from the National Productivity Investment Fund (MR/R026408/1). SMK and JFW work in a unit

that receives support from the UK Medical Research Council (MC_UU_00007/10). DIB receives support from a Royal Netherlands Academy of Science Professor Award (PAH/6635). DJB receives support from NIA/NIH grants R24-AG065184 and R01-AG042568 to UCLA and R56-AG058726 to the University of Southern California, Open Philanthropy (010623-00001), Ragnar Söderberg Foundation (E42/15), the Swedish Research Council (421-2013-1061, 2019-00244). MB is funded by an ERC Consolidator Grant (WELL-BEING; grant 771057). JBP is supported by the Medical Research Foundation 2018 Emerging Leaders 1st Prize in Adolescent Mental Health (MRF-160-0002-ELP-PINGA) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 863981). WDH is supported by Age UK (Disconnected Mind grant), and by a Career Development Award from the Medical Research Council (MRC) [MR/T030852/1] for the project titled "From genetic sequence to phenotypic consequence: genetic and environmental links between cognitive ability, socioeconomic position, and health". DME is supported by an NHMRC Senior Research Fellowship (GNT1137714). JLH is a NHMRC Senior Principal Research Fellow. SL is supported by a Victorian Cancer Agency Early Career Research Fellowhip (ECRF19020). MCS is a Senior Research Fellow of the National Health and Medicial Research Council of Australia. SEM and NGM are supported through NHMRC investigator grants (APP1172917 and APP1172990). BMB, BOÅ, HR, AFH and KH work in a research unit funded by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU; The Liaison Committee for education, research and innovation in Central Norway; the Joint Research Committee between St. Olavs Hospital and the Faculty of Medicine and Health Sciences, NTNU. ON receives support from own institution and Norwegian Research Council grant number 287347. AH was supported by a career grant from the South-Eastern Norway Regional Health Authority (2020022). MCM is supported by the Leverhulme Trust, Leverhulme Centre for Demographic Science, ERC 835079. CH is supported by an MRC University Unit Programme Grant MC_UU_00007/10 (QTL in Health and Disease). JKH was supported by DA011015 for collection of the data used in this paper, and is currently supported by U01DA051018, R01DA042755, and R01AG046938. MCK is supported by National Institute of Mental Health grant R01 MH100141.

Contributions

LJH, MGN, TTM, YC, JBP, JFW, JLH, SL, MCS, DAL, NGM, AH, KH, CJW, BOÅ, PDK, JK, SEM, DJB, PT, DME, GDS, CH, BMB, GH and NMD were closely involved in conceptualising and designing the study.

JK, KPH, EMTD, SMK, HC, JFW, EJCD, RP, JAS, PAP, SLRK, SL, JLH, MCS, KC, NMD, SEM, NGM, BMB, RGW, IYM, KL, KH, CJW, CRB, AEJ, DP, CH, AC were involved in data and funding acquisition.

LJH developed the GitHub GWAS pipeline with support from GH and NMD and programming code from GH and PT (via SSGAC). CH kindly beta tested the GWAS pipeline and suggested improvements. Other analysts (listed below) also made major contributions to the GWAS pipeline.

LJH, SG, AFH, HR, CH, YC, GC, PAL, TP, MVDZ, RC, MM, YW, SL, LK, SMR, LFB, CAR, MN, JVB performed GWAS analyses in individual cohorts with the support and guidance of NMD, GH, BMB, RGW, IYM, KL, AEJ, SEM, JK, MGN, MB, JBP, SH, JLH, JFW, JAS, PAP, SLRK, KC, MCK.

LJH performed meta-analyses and all downstream analyses with the meta-analysis data.

LJH drafted the first version of the manuscript. MGN, TTM BOÅ, PDK, JK, SEM, RGW, DJB, PT, DME, GDS, CH, BMB, GH and NMD played a key role in interpreting the results, planning additional analyses, and revising the manuscript.

All authors contributed to and critically reviewed the manuscript.

Data availability

GWAS summary statistics will be made publicly available for download prior to publication.