nature genetics

Genotype–covariate interaction effects and the heritability of adult body mass index

Matthew R Robinson[1,2], Geoffrey English[1], Gerhard Moser[1], Luke R Lloyd-Jones[1], Marcus A Triplett[3], Zhihong Zhu[1], Ilja M Nolte[4], Jana V van Vliet-Ostaptchouk[4,5], Harold Snieder[4], The LifeLines Cohort Study[6], Tonu Esko[7–10], Lili Milani[7], Reedik Mägi[7], Andres Metspalu[7,11], Patrik K E Magnusson[12], Nancy L Pedersen[12], Erik Ingelsson[13,14], Magnus Johannesson[15], Jian Yang[1], David Cesarini[16] & Peter M Visscher[1]

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

Obesity is a worldwide epidemic, with major health and economic costs. Here we estimate heritability for body mass index (BMI) in 172,000 sibling pairs and 150,832 unrelated individuals and explore the contribution of genotype–covariate interaction effects at common SNP loci. We find evidence for genotype–age interaction (likelihood ratio test (LRT) = 73.58, degrees of freedom (df) = 1, $P = 4.83 \times 10^{-18}$), which contributed 8.1% (1.4% s.e.) to BMI variation. Across eight self-reported lifestyle factors, including diet and exercise, we find genotype–environment interaction only for smoking behavior (LRT = 19.70, $P = 5.03 \times 10^{-5}$ and LRT = 30.80, $P = 1.42 \times 10^{-8}$), which contributed 4.0% (0.8% s.e.) to BMI variation. Bayesian association analysis suggests that BMI is highly polygenic, with 75% of the SNP heritability attributable to loci that each explain <0.01% of the phenotypic variance. Our findings imply that substantially larger sample sizes across ages and lifestyles are required to understand the full genetic architecture of BMI.

A fundamental question in biology is the degree to which observed variation in phenotype is due to environmental or genetic factors. Heritability (h^2) is a simple dimensionless population-specific parameter of the proportion of phenotypic variation that is attributable to genetic factors[1,2]. Analogously, in populations of traditionally unrelated individuals, the SNP heritability (h^2_{SNP}) is a population-level parameter of the proportion of phenotypic variation attributable to a set of SNP markers. In human populations, these parameters quantify how much of the resemblance between relatives is due to shared genetic factors (as opposed to shared environmental factors)[1,2], allow a comparison of traits within and across populations[3], determine the efficiency of gene-mapping studies[4], and provide the upper bound for prediction of the genetic risk of disease[5].

Estimates of the heritability of BMI differ substantially across experimental designs (**Table 1**). Current literature suggests that family studies designed to have fewer biases and confounders result in estimated BMI h^2 values of ~0.4 (refs. 6–12), around half that from classical twin studies (between 0.6 and 0.8)[7,13] (**Table 1**). In a population design, an estimate of h^2_{SNP} in distant relatives is unbiased by shared environment, as distant relatives are unlikely to share variation due to environmental factors[14,15]. Although not all additive genetic variance is captured by common SNPs, a comparison of h^2_{SNP} across traits can be informative. Studies have estimated h^2_{SNP} at common HapMap3 loci to be 0.22 (s.e. = 0.02) for BMI[16], approximately half that of height for the same data[17] (0.50, s.e. = 0.04). Recently, a study using genotypes imputed to a sequence reference and a design that fits multiple relationship matrices estimated from a range of SNP sets stratified by linkage disequilibrium (LD) and minor allele frequency (MAF) estimated h^2_{SNP} of BMI (0.27, s.e. = 0.03) to be half that for height (0.56, s.e. = 0.02) and found no evidence that differences in h^2_{SNP} between the traits are explained by a greater contribution of untagged rare variants to BMI[15]. Additionally, previous studies suggest that dominance effects have little role in creating variation in height or BMI[18]. Taken together, population studies imply that estimates of h^2 for BMI are systematically inflated in classical twin studies.

The systematic inflation of estimated h^2 values for BMI in classical twin studies may simply reflect confounding between shared environment and genetic effects. Alternatively, shared age and environments between twins could contribute to the discrepancy in estimates across experimental designs, if there are strong age- or environment-dependent genetic effects. If this is the case, heritability estimates from population studies of unrelated individuals might represent the average

[1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia. [2]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [3]Queensland Brain Institute, The University of Queensland, Brisbane, Australia. [4]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [5]Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [6]A full list of members and affiliations appears in the **Supplementary Note**. [7]Estonian Genome Center, University of Tartu, Tartu, Estonia. [8]Division of Endocrinology, Boston Children's Hospital, Cambridge, Massachusetts, USA. [9]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA. [10]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [11]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [12]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [13]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [14]Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, California, USA. [15]Stockholm School of Economics, Stockholm, Sweden. [16]Center for Experimental Social Science, Department of Economics, New York University, New York, New York, USA. Correspondence should be addressed to M.R.R. (matthew.robinson@unil.ch) or P.M.V. (peter.visscher@uq.edu.au).

Received 5 July 2016; accepted 12 June 2017; published online 10 July 2017; doi:10.1038/ng.3912

Table 1 Overview of study designs estimating heritability of BMI

Study design	Heritability estimate (s.e.)	Benefits	Pitfalls	Reference
Classical twin design[a]	0.63 (0.01) 0.75 (0.02)	High precision	Limited by two correlation estimates Environmental sharing among different relatives difficult to separate Confounding by nonadditive sources of genetic variance	8,14,15
Family studies[b]				
Siblings raised together or apart	0.44 (0.04)	High precision Low bias by environmental sharing	Large sample sizes required	This study
Pedigree estimates	0.46 (0.06) 0.39 (0.04)	High precision	No separation of environmental sharing from genetic similarity	15,16
Within-family segregation[c]	0.42 (0.02) 0.42 (0.17)	No confounding by population stratification Not reliant on assumptions of environmental sharing	Low precision due to high standard errors Large sample sizes required	12,13,18
Population studies[d]				
HapMap3 SNPs	0.22 (0.02)	Unrelated individuals unlikely to share environments Unbiased by nonadditive genetic effects	Captures only variance attributable to loci in LD with genetic markers	17,20
Whole-genome imputation	0.27 (0.03)	Unrelated individuals unlikely to share environments Unbiased by nonadditive genetic effects Imputation provides better genome coverage	Captures only variance attributable to loci in LD with genetic markers	19

[a]Classical twin studies estimate heritability using the observed and expected resemblance between twin pairs. [b]In family studies, heritability is estimated from familial (pedigree) data, by modeling the observed and expected resemblance between siblings or across wider-degree relatives (pedigree estimates). [c]Studies of within-family segregation avoid dependence on modeling assumptions about between-family variance by estimating heritability through correlating phenotypic similarity with realized genetic sharing estimated from genomic markers. [d]Population studies estimate the proportion of phenotypic variance attributable to genomic markers (SNP heritability).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

genetic effects across ages and environmental or lifestyle factors, resulting in unobserved genetic variation. Here we use a large body of data on siblings and publicly available phenotype–genotype data sets of unrelated individuals to estimate the heritability of BMI across a range of experimental designs and explore the contribution of age- and environment-specific genetic effects to variation in BMI.

RESULTS

The heritability of BMI from siblings

We first used a combined sample of 172,000 pairs of 18-year-old, male full and half-siblings[19,20], in an experimental design that is likely to have few biases and confounders, to estimate heritability of BMI across a range of relative pairs. The estimated h^2 of BMI declined from 0.770 (s.e. = 0.012) for monozygotic pairs and 0.637 (s.e. = 0.014) for dizygotic pairs in a classical twin design to 0.557 (s.e. = 0.057) for full siblings raised apart and 0.436 (s.e. = 0.037) for half-siblings raised apart (**Fig. 1**, **Supplementary Tables 1** and **2**, and **Supplementary Note**). In contrast, h^2 estimates for height were 0.853 (s.e. = 0.012) for monozygotic twin pairs, 0.762 (s.e. = 0.010) for dizygotic twin pairs, and 0.763 (s.e. = 0.036) for half-siblings raised apart (**Fig. 1** and **Supplementary Table 1**). These patterns are consistent with differences in phenotypic correlation of BMI and height between both full and half-siblings raised together and apart (**Supplementary Table 1**) and between paternal and maternal half-siblings (**Supplementary Table 2**). The estimate obtained from half-siblings raised apart is similar to that from previous studies of within-family segregation[10,12], which avoided dependence on modeling assumptions regarding between-family variance by using realized genetic sharing (**Table 1**). Taken together, these results support a h^2 for BMI of ~0.4 (refs. 6–12) and a systematic inflation of BMI heritability estimates in classical twin studies. We now explore whether genotype–age and genotype–environment effects could contribute to this inflation or whether it is due simply to confounding between relatedness and the degree of shared developmental environment.

Genotype–covariate interactions in population studies

We present a framework for analyzing genome-wide gene–covariate interaction effects within a population sample. We define genotype–covariate interaction as a change in SNP marker effects across a covariate that alters the proportion of phenotypic variance attributable to the SNPs and/or results in a genetic correlation of <1 (**Fig. 2**). Throughout, we assume that a phenotype has been measured at t time points and that a covariate of interest changes with t. At each time point, we adjust for age and sex effects and standardize the phenotype to mean 0 and variance 1 with a rank inverse normal transformation to correct for mean–variance relationships and adjust for heterogeneity in phenotypic variance. With an inverse normal transformation, genotype–covariate interaction is detected if changes in genetic effects do not scale proportionally with changes in mean or phenotypic variance across ages (i.e., there is a change in h^2_{SNP} and/or the correlation of the SNP effects is <1). We repeated our analyses without transformation at each time point to examine heterogeneity of variance and to confirm that changes in h^2_{SNP} across time points reflect a change in variance tagged by SNPs rather than a change in environmental effects (residual variance) (**Supplementary Note** and **Supplementary Fig. 1**). Our combined approach: (i) estimates the variance explained by SNPs and the covariance among SNP effects across different points of measurement (along a continuous gradient or between factor levels), (ii) tests for gene–covariate interaction, and (iii) estimates the proportion of phenotypic variance contributed by genotype–covariate interaction effects, all of which are required to fully describe gene–covariate interaction (**Fig. 2**). The models we used, the parameters estimated, and the hypotheses we tested are described in **Supplementary Table 3**.

SNP heritability and the correlation of SNP effects across a covariate. We first estimated a full multitrait restricted maximum likelihood model (MV-GREML) that estimates h^2_{SNP} at each point t and the genetic correlations among measurement points[21]. We tested whether

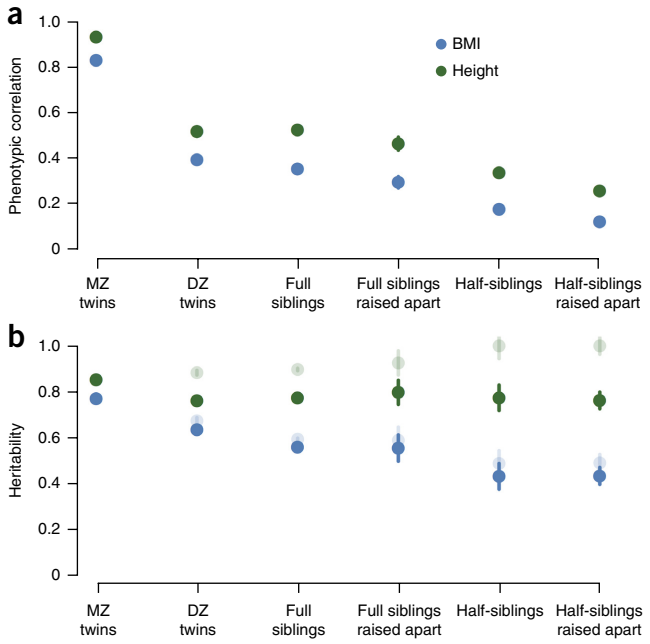

Figure 1 Systematic inflation of BMI heritability estimates in close relatives that share developmental environments. (**a**,**b**) Phenotypic correlations (**a**) and heritability estimates (**b**) from behavioral genetic models (see Online Methods) among different male sibling pairs taken from Swedish army conscription BMI (blue) and height (green) records from 1950 to 1969. In **b**, estimates are presented assuming assortative mating for both traits (see Online Methods), with transparent points giving the estimate when trait assortment is ignored or assumed absent. Error bars, s.e.

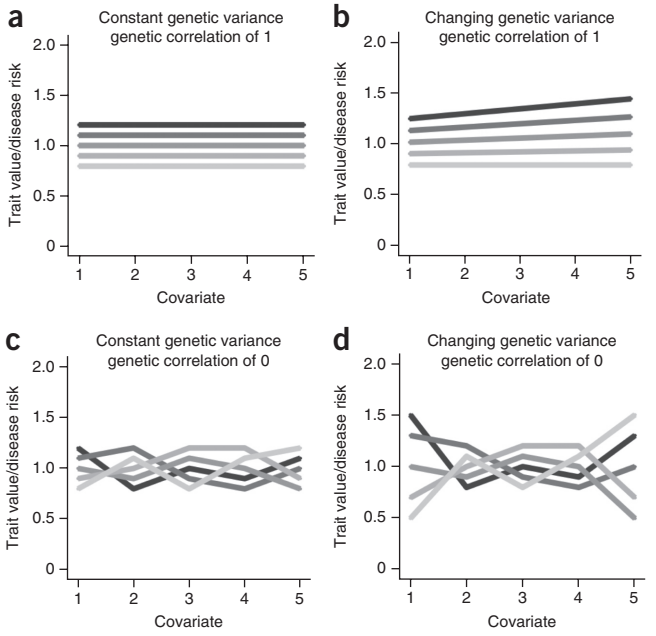

Figure 2 Genotype–covariate interactions described through estimation of genetic variance and genetic correlations across a gradient. (**a**–**d**) Different scenarios of genotype–covariate interaction with shaded lines giving expected phenotypic values for five genotypes across five measurement points. (**a**) When there is no genotype–covariate interaction, differences among genotypes are constant across the covariate and the correlation of SNP marker effects (genetic correlation) = 1 across measurements. (**b**) Changing differences among genotypes with a constant ordering results in increased genetic variance along the covariate but a genetic correlation of 1 among measurements. (**c**) Changing ordering of genotypes, with a genetic correlation of 0 across all pairs of measurement points but constant genetic variance. (**d**) Changing ordering of genotypes, with a genetic correlation of 0 across all pairs of measurement points, and differences in genetic variance.

this model provides a better fit to the data than a GREML model with one genetic and one residual component by LRT (Online Methods). This test provides evidence for heterogeneity of the variance components across t measurement points (**Supplementary Table 3** and **Supplementary Note**).

Testing for genotype–covariate interaction. Next, through the use of a covariance function approach, we reanalyzed the data by modeling the variance and covariance of genome-wide SNP effects across t measurement points with a reduced number of parameters ('random regression' REML model (RR-GREML))[22–25]. This approach provides a test for the presence of gene–covariate interaction by comparing the fit to the data from a RR-GREML model, where SNP effects differ across the covariate (through a first-order polynomial function, $k = 1$), to a RR-GREML model, where the SNP effects are constant across a covariate (zero-order polynomial function, $k = 0$), using LRT (Online Methods and **Supplementary Table 3**). This approach has been developed for repeated measures data of close relatives in animal breeding and evolutionary biology[22–26], and here we applied it to population studies where unrelated individuals were measured at t points along a gradient, with each individual having only a single measurement. Repeated measures across life are not required in this setting, as estimated genetic relationships among conventionally unrelated individuals measured at each point are used, resulting in inference that is unbiased by confounders such as common or shared environment, because phenotypes are measured on different people across t. Theory and the results of a simulation study indicate that MV-GREML and RR-GREML with a polynomial of $k = t - 1$ are equivalent models and that a RR-GREML approach provides an appropriate testing framework for gene–covariate interactions in population studies of conventionally unrelated individuals (see Online Methods, **Supplementary Note** and **Supplementary Fig. 1**).

Phenotypic variance contributed by interactive effects. Third, we estimated the proportion of phenotypic variance contributed by genotype–covariate interaction effects using a model proposed by Yang *et al.*[27]. We used this approach to estimate the coheritability of phenotype across t, estimate the proportion of variance attributable to genotype–covariate interaction effects (GCI-GREML model; Online Methods), and test for the presence of gene–covariate interaction by comparing the fit to the data of a GCI-GREML model (fitting a genetic variance component and a genetic interaction variance component) and a null GREML model (fitting a single genetic variance component) by LRT (**Supplementary Table 3**). For each covariate, we also used GCI-GREML models to test for evidence of genotype–covariate interaction effects between pairs of measurement points. Theory and the results of a simulation study indicate that a GCI-GREML approach provides an appropriate testing framework for gene–covariate interactions and estimates the variance contributed by genotype–covariate interaction effects (see Online Methods, **Supplementary Note** and **Supplementary Fig. 2**).

In our analyses of genotype–age and genotype–environment interaction effects, we conducted a total of 250 LRTs (Online Methods) and thus adopted a Bonferroni multiple-testing threshold, with $P = 2.00 \times 10^{-4}$.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

Table 2 Testing for genotype–environment interactions for BMI in 97,510 participants of the UK Biobank study with MV-GREML, RR-GREML, and GCI-GREML models

Covariate	t	logLK MV-GREML	LRT MV-GREML versus GREML	logLK RR-GREML $k = 0$	LRT RR-GREML $k = 0$ versus GREML	logLK RR-GREML $k = 1$	LRT RR-GREML $k = 1$ versus RR-GREML $k = 0$	logLK GCI-GREML	LRT GCI-GREML versus GREML	AIC MV-GREML	AIC RR-GREML $k = 0$	AIC RR-GREML $k = 1$	AIC GCI-GREML
Diet change	3	48.05	6.87×10^{-18}	37.52	5.07×10^{-17}	45.73	2.72×10^{-4}	0.00	0.500	94,525.04	94,536.10	94,523.68	94,609.14
Diet variation	3	34.93	1.58×10^{-12}	32.08	1.17×10^{-14}	32.93	0.427	0.00	0.500	94,557.90	94,553.60	94,555.90	94,615.76
Self-reported health	3	43.95	3.34×10^{-16}	35.59	3.50×10^{-16}	42.77	7.62×10^{-4}	0.02	0.421	90,676.48	90,683.20	90,672.84	90,752.34
Pack years of smoking	4	380.82	2.77×10^{-155}	82.22	2.02×10^{-35}	92.07	5.27×10^{-5}	15.40	1.43×10^{-8}	93,839.80	94,419.00	94,403.30	94,548.64
Weekly periods of exercise	9	170.58	3.93×10^{-44}	147.04	7.49×10^{-59}	148.15	0.330	0.25	0.240	94,419.13	94,378.22	94,380.00	94,657.80
Weekly alcohol consumption	8	151.78	2.41×10^{-41}	130.11	1.85×10^{-52}	135.45	0.005	1.15	0.065	94,427.58	94,400.92	94,394.24	94,646.84
Weekly hours of television	6	95.92	1.13×10^{-27}	88.04	3.68×10^{-36}	89.07	0.357	0.89	0.091	94,465.28	94,441.04	94,442.97	94,607.34
Local poverty index (TDI)	10	183.51	1.80×10^{-44}	162.22	1.69×10^{-64}	162.63	0.664	0.00	0.500	94,372.92	94,309.51	94,312.68	94,617.94

The log-likelihood values (logLK) for each model are given as a difference from a GREML model that estimates a single genetic and residual component. The P values from the LRT statistics are given for the model comparisons as twice the difference in logLK between a MV-GREML model and a GREML model with $(t(t + 1)/2 + t) - 2$ df; a RR-GREML model of $k = 0$ and a GREML model with $(t + 1) - 2$ df; a RR-GREML model of $k = 1$ and a RR-GREML model of $k = 0$, and a GCI-GREML model and a GREML model with 1 df (one tailed). AIC, Akaike information criterion.

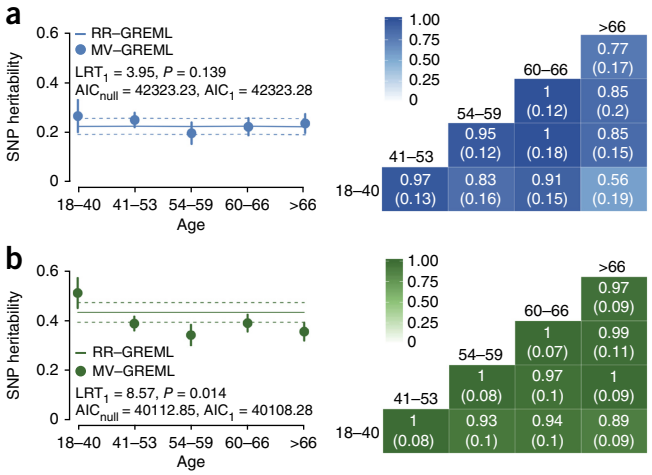

Figure 3 Genotype–age interactions for BMI and height in 43,407 individuals in the composite AHTHEL sample. (**a**,**b**) MV-GREML and RR-GREML estimates of the proportion of phenotypic variation attributable to common HapMap3 SNP markers (left) and of the correlation of SNP effects (right) across five age groups for BMI (**a**) and height (**b**). Akaike information criterion (AIC) values are also shown. Zero-order RR-GREML estimates are shown in **a** and **b**. Dashed lines and numbers in parentheses indicate s.e. (Online Methods). Age groups (in years) and numbers of individuals (n) in each were as follows: 18–39 (n = 5,136) 40–53 (n = 11,573); 54–59 (n = 7,829) 60–66 (n = 9,843), and ≥67 (n = 9,025).

The heritability of BMI across age

Longitudinal twin and extended family studies have examined genetic correlations across different ages for several complex traits, including BMI[28,29] and height[30], finding little evidence for changes in heritability with age. In population studies, a meta-analysis of 320,485 individuals tested for age-specific genetic effects and identified 15 loci with age-specific effects on BMI, of which 11 had larger effects in people younger than 50 years old[31]. An estimate of $h^2_{SNP} = 0.21$ (s.e. = 0.03) for BMI was reported, with no significant difference between age groups[31], suggesting that individual loci may differ in the strength of their association with BMI across life span, but that these effects are not likely to contribute substantially to the phenotypic variance.

We estimated the genetic (co)variance of BMI captured by common HapMap3 SNP loci across adult life span in a large composite data set of 43,407 individuals measured between 18 and 80 years of age (AHTHEL composite cohort data, **Supplementary Table 4**). We divided the sample into five age groups, representative of different stages of adult life. We found no evidence for change in h^2_{SNP} for BMI across adult life span, and a first-order polynomial RR-GREML model did not provide a better fit to the data than a zero-order polynomial RR-GREML model (LRT = 3.95, $P = 0.139$; **Fig. 3**). However, when using a GCI-GREML model, we found evidence for significant genotype–age interaction (LRT = 73.58, $P = 4.83 \times 10^{-18}$), with a SNP coheritability estimate of 21.1% (s.e. = 0.9%), and 8.1% (s.e. = 1.4%) of BMI variance attributable to genotype–age interaction effects within this sample. This result was driven by a difference in SNP effects between young and old individuals. When we analyzed only young (18–40 years) and old (>66 years) individuals together (GCI-GREML model, n = 14,106), the proportion of BMI variance attributable to genotype–age interaction effects tagged by the SNPs was 11.7% (s.e. = 2.5%) within this sample (LRT = 27.71, $P = 7.05 \times 10^{-8}$). We found similar results without inverse normal transformation (SNP coheritability estimate: 0.210 (s.e. = 0.010), estimate of variance of

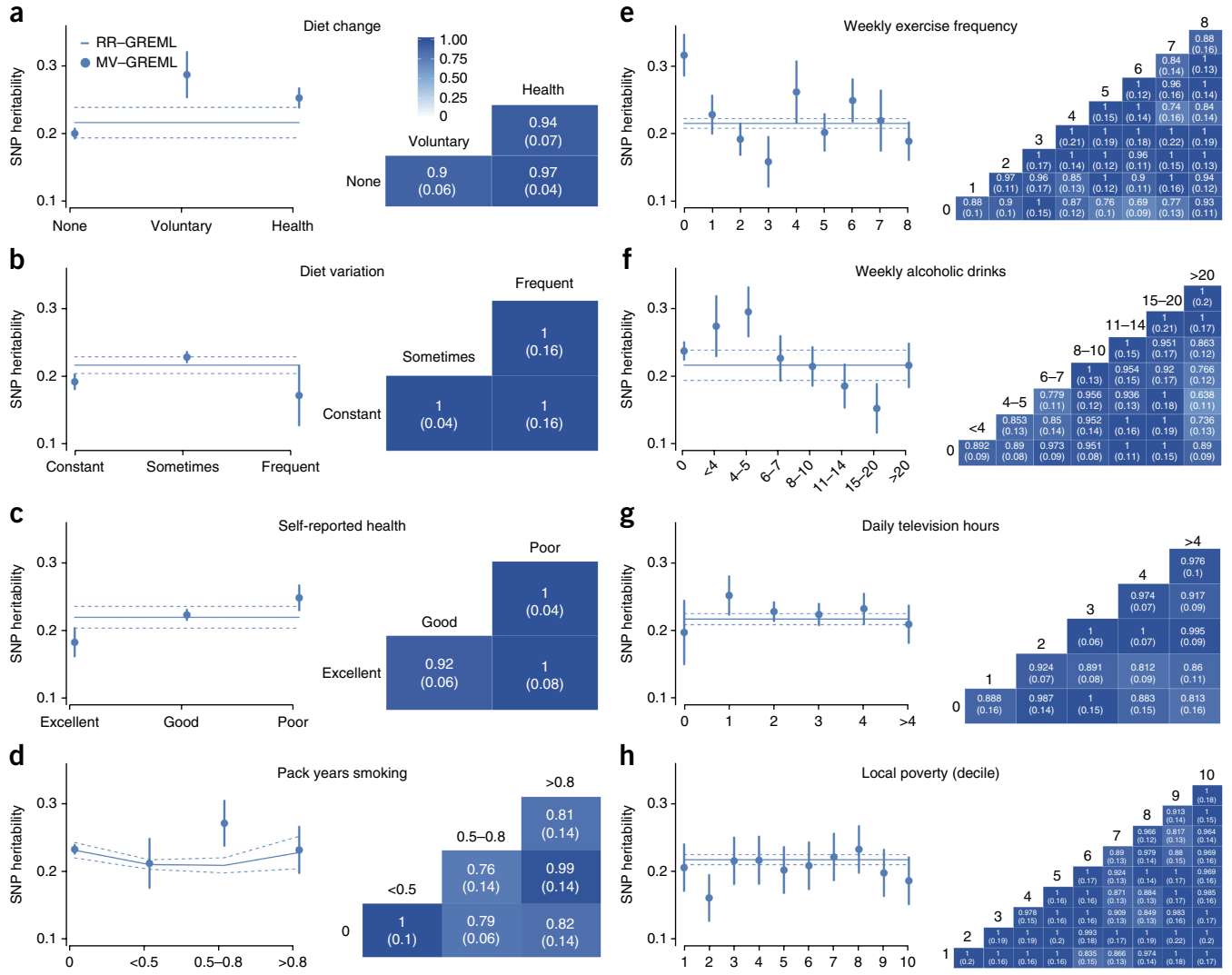

Figure 4 Genotype–environment interactions for BMI in 97,510 participants of the UK Biobank study across a range of lifestyle factors. (**a**–**h**) Estimates of the phenotypic variance captured by common HapMap3 SNP loci (SNP heritability) from MV-GREML models (error bars, s.e.) and RR-GREML models (dashed lines show s.e.) for each lifestyle factor. For the RR-GREML model estimates, we present the model of best fit to the data, as assessed by LRT (**Table 2**). Also shown are estimates of the correlation of SNP marker effects (genetic correlation) across groups for each lifestyle factor (s.e. given in parentheses).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

genotype–age interaction effects tagged by the SNPs: 0.081 (s.e. = 0.015), LRT = 71.15, $P = 1.61 \times 10^{-17}$), confirming that changes in genetic effect for BMI are not created simply by a change in environmental effects (residual variance) across time points but reflect changing SNP marker effects with age.

For comparison, we repeated our analyses for height in the AHTHEL data. We found no evidence for genotype–age interaction effects with inverse normal transformation (LRT GCI-GREML model = 0.132, $P = 0.358$, SNP coheritability estimate: 0.439, s.e. = 0.010, estimate of variance of genotype–age interaction effects tagged by the SNPs: 0.006, s.e. = 0.013) (**Fig. 3**) or without (LRT GCI-GREML model = 0.28, $P = 0.298$, SNP coheritability estimate: 0.436, s.e. = 0.011, estimate of variance of genotype–age interaction effects tagged by the SNPs: 0.006, s.e. = 0.012, LRT = 0.28, $P = 0.298$) (**Supplementary Fig. 3**).

We repeated our analysis of BMI for individuals aged 46–73 in the UK Biobank study and found no evidence for genotype–age interaction effects (**Supplementary Fig. 4**) (LRT RR-GREML model = 1.02, $P = 0.601$, LRT GCI-GREML model = 0.846, $P = 0.179$, $n = 107{,}488$, estimate of variance of genotype–age interaction effects tagged by the SNPs: 0.001, s.e. = 0.005), supporting our findings in the AHTHEL sample in these age groups (GCI-GREML model of ages 41–53 and >66 in the AHTHEL sample, $n = 20{,}599$, LRT = 1.30, $P = 0.128$).

The heritability of BMI across environments

Environmental factors such as diet, exercise, and lifestyle influence BMI, and the risk of obesity varies across different human populations[32]. It is frequently proposed that these environmental factors interact with genetic predisposition for BMI[33]. Studies of genotype–environment interaction for BMI using a cross-sectional approach have found that the effects of a single environmental variable (e.g., diet[34] or lifestyle[35]) may depend on a specific genetic polymorphism (for example, SNPs in the *FTO* gene)[36–38] or a genetic predictor created from genome-wide significant loci[39]. However, these detected effects typically explain little variation, and the contribution of genotype–environment interaction effects to the phenotypic variance of BMI within a population has not been fully quantified.

Table 3 GCI-GREML estimation of genetic interaction variance for BMI in 97,510 participants of the UK Biobank study

Covariate	V(G)	V(GCI)	V(GCI) extremes	LRT GCI-GREML extremes versus GREML	LRT *P* value
Diet change	0.217 (0.006)	1.02×10^{-6} (0.007)	0.007 (0.009)	0.50	0.240
Diet variation	0.215 (0.006)	0.001 (0.007)	1.00×10^{-5} (0.010)	0.00	0.500
Self-reported health	0.218 (0.006)	0.002 (0.007)	0.029 (0.018)	2.77	0.048
Pack years smoking	0.196 (0.006)	0.040 (0.008)	0.042 (0.006)	19.20	5.89×10^{-6}
Weekly periods of exercise	0.215 (0.005)	0.007 (0.009)	0.015 (0.027)	0.30	0.292
Weekly alcohol consumption	0.214 (0.005)	0.014 (0.009)	0.040 (0.018)	5.03	0.012
Weekly hours of television	0.215 (0.005)	0.011 (0.008)	0.053 (0.034)	2.51	0.057
Local poverty index (TDI)	0.216 (0.005)	1.00×10^{-6} (0.011)	0.003 (0.034)	0.01	0.465
Total		0.075	0.189		

V(G), the coheritability estimate of the trait across environmental measures of the trait; V(GCI), estimate of the variance attributable to genotype–covariate interactive effects. We repeated the testing using only data from the extremes and present the estimate of V(GCI), the LRT statistics comparing a GCI-GREML model and a GREML model of the extremes, and the associated *P* value at 1 df.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

Using a series of eight self-report variables of diet, exercise, and lifestyle shown to explain a combined 14% of the phenotypic variation of age and sex-adjusted BMI in the UK Biobank sample (n = 97,510; **Supplementary Table 5** and **Supplementary Fig. 5**), we found evidence that a MV-GREML model provides a better fit to the data than a GREML model for all eight covariates (**Table 2**). This suggests heterogeneity of the variance components across these environmental variables. We then tested whether this heterogeneity reflects genotype–environment interactions for BMI and found evidence for genetic interaction effects with smoking behavior (LRT = 19.70, $P = 5.03 \times 10^{-5}$ from a RR-GREML model) (**Table 2** and **Fig. 4**) that make a significant contribution of 4.0% (s.e. = 0.8%) to the phenotypic variance (LRT = 30.80, $P = 1.42 \times 10^{-8}$ from a GCI-GREML model) (**Table 3**). The sum across environmental covariates of the proportions of variance attributable to genotype–covariate interaction effects estimated in the UK Biobank sample was 7.5% (range 1.00×10^{-4}% to 4.0%) (**Table 3**).

We repeated our analysis without inverse normal transformation and found widespread heterogeneity in variance, with higher phenotypic variance and variance tagged by common SNP loci for individuals who did not exercise, changed their diet for health reasons, varied their diet, watched a large amount of television, reported poor health, smoked heavily, did not report extreme alcohol consumption, and resided in low socioeconomic areas (**Supplementary Fig. 6** and **Supplementary Table 6**). These conditions promote higher BMI on average within the population (**Supplementary Table 5**); therefore, these effects probably represent mean–variance scaling[40], and, as we show above, when we correct for mean–variance relationships and adjust for heterogeneity in phenotypic variance, only smoking behavior is found to influence SNP marker effects for BMI.

Gene-discovery studies of BMI

By itself, heritability is not informative about the number of causal loci or their effect sizes. Power is proportional to the amount of phenotypic variation explained by a variant, so if we assume the same number of genetic variants for both BMI and height, the variance explained is halved if the heritability is ~50% lower for BMI, implying that twice the sample size is required for detection. This may explain why, in the most recent genome-wide association study (GWAS) of BMI[16] and height[17], ~7-fold more variants were found to be robustly associated with height, even though larger sample sizes were used for BMI. Here we investigate this further using a recently proposed Bayesian mixture model that provides inference on genetic architecture (Online Methods)[41]. Although the true effect size distribution and the pattern of correlation between unobserved causal variants and genotyped SNPs is unknown, a comparison of SNP numbers, effects, and variance explained across phenotypes for which the genotyped SNPs are the same enables a comparison of the distribution of effect sizes at causal variants tagged by those SNPs.

We applied this model to BMI and height in 107,488 individuals of European ethnicity in the UK Biobank study. We found that 75% of the h^2_{SNP} of BMI is attributed to loci that each explain <0.01% of the variance (**Fig. 5**). In contrast, 50% of the h^2_{SNP} of height was attributed to loci that each contribute >0.01% of the variance (**Fig. 5**). Therefore, more variation in BMI can be attributed to common loci of very small effect than to height, implying that BMI is likely more polygenic. With sample sizes approaching 1 million people in the near future, power will be 0.99 for loci that contribute 0.01%. However, loci with effects ≥0.01% cumulatively explain only ~25% of h^2_{SNP} for BMI. For the remaining 75%, power at a sample size of 1 million is 0.95 for loci contributing 0.005% but only 0.01 for those contributing 0.001%, suggesting that smaller effect size loci may remain elusive in GWAS of BMI because many loci have far smaller effects than those for height.

DISCUSSION

Our findings are consistent with a narrow-sense heritability of BMI of ~40% and suggest that most BMI-associated loci are likely to explain less phenotypic variance than those for height, with the notable exception of *FTO*. It is often suggested that genotype–age and genotype–environment interaction effects are important for BMI, and here we found evidence for genotype–age interaction effects and genotype–environment interaction effects with smoking behavior, which explain 8.1% (s.e. = 1.4%) and 4.0% (s.e. = 0.8%) of the phenotypic variance of BMI, respectively. Self-report measures of environmental covariates may be inaccurate, and potentially important variables may not be recorded at all, meaning that there are probably additional undetected covariates for which genotype–covariate interaction occurs. However, our results imply that additional undetected genotype–covariate interactions may each explain very little variation at the population level, contrary to previous studies[34–37,39]. First, with the exception of smoking behavior, we find no evidence of genotype–environment interaction at a set of environmental variables that explain a large amount (14%) of the phenotypic variation of age- and sex-adjusted BMI, with point estimates of the proportion of variance attributable to interactive effects of 1% or less for each factor (**Table 3**). Second, for variables such as age and recall of smoking behavior, which are likely to be accurately recorded and have large effects on

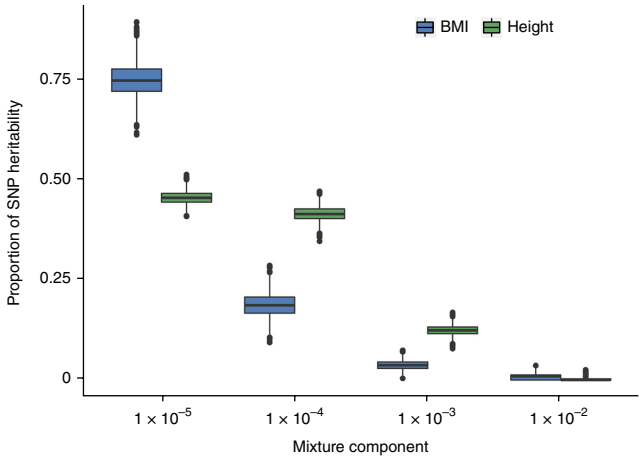

Figure 5 A Bayesian mixture model of common HapMap3 SNP markers for BMI and height in 107,488 participants of the UK Biobank study. BayesR model estimates of the proportion of genetic variance contributed by SNPs with different mixture distributions of effect sizes for BMI (blue) and height (green) are shown; error bars, 95% credible intervals of the posterior distribution.

mean BMI, the genotype–covariate interaction effects we detected explain limited amounts of phenotypic variance. Therefore, although testing for genotype–covariate interaction effects at additional and improved measures of diet and exercise remains to be conducted, our results provide an expectation of small effect sizes.

When using a GCI-GREML model to test for the presence of genotype–covariate interaction among groups, if we assume two groups each with h^2_{SNP} for BMI of 0.25, a genetic correlation of 0.8, and equal sample size of 9,751 per group (97,510 UK Biobank individuals divided in decile groups), giving an expected proportion of variance attributable to genotype–covariate interaction effects of 0.05 (**Supplementary Note**), then our power to detect an interaction at our multiple testing threshold of 2.00×10^{-4} is 50%. If h^2_{SNP} were 0.3 and 0.2 in the two groups, with all other parameters being the same (5.4% phenotypic variance attributable to genotype–covariate interaction effects), our power would increase to 60%, and to 96% if the genetic correlation were 0.7 (or if 7.9% of the phenotypic variance were attributable to genotype–covariate interaction effects). However, when estimating a GCI-GREML model on the full data of 97,510 individuals, power approaches 100% if 5% phenotypic variance is attributable to genotype–covariate interaction effects and is 96% at 4%, and 65% at 3% variance explained. Therefore, given the number of tests conducted, power in this study is high to detect important genome-wide genotype–covariate interaction effects that explain 4% or more of the phenotypic variance, but substantially larger sample sizes will be required to detect interactive effects that explain small amounts of phenotypic variance.

Other factors may also influence h^2 of BMI across study designs. In both the pedigree and the within-family design, nonadditive genetic variation may contribute to the heritability estimate[2]. We tested examined dominance variance for BMI within the UK Biobank sample and found no evidence of this ($d^2_{SNP} = 0.002$, s.e. = 0.004, $P = 0.492$), in line with a recent study[18]. Additionally, rare variants may have a greater influence on BMI variation, but a recent study found that imputed rare variants explain only ~4% of the variance[15]. Furthermore, assortative mating creates a correlation at height- and BMI-associated loci among couples[42], which may inflate h^2 across study designs, but it is not clear whether mate choice is historically consistent or whether equilibrium has been reached. Finally, the contribution of autosomal sex-specific genetic effects to variation in BMI is a longstanding question[43,44]; however, although there is mixed evidence from classical twin studies for a cross-sex correlation less than unity[6,13], gene-discovery studies[45] and population studies[31,46,47] have found no evidence for sex differences in h^2, sex-specific effects at autosomal genetic variants, or a cross-sex genetic correlation less than unity. Therefore, we suggest that genotype–age and genotype–environment effects may contribute to the inflation of BMI heritability estimates in classical twin studies, in combination with stronger common environment effects in close relatives than more distant ones.

Future research could focus on early-life obesity and the maternal effects that may contribute to variance and improve understanding of the biological mechanisms that promote disease in early adult life. The framework we present here can be used to study any form of genotype–covariate interactive effects across binary or continuous variables, such as sex, population, or early-life conditions. In this case, we show that untangling the full genetic architecture of BMI via GWAS will require substantially larger sample sizes across ages and lifestyle factors than are currently available and a better understanding of the effects that contribute to between-person variability in obesity.

URLs. UK Biobank documentation, http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf and http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/imputation_documentation_May2015.pdf; GCTA, http://cnsgenomics.com/software/gcta/; HapMap3, ftp://ftp.ncbi.nlm.nih.gov/hapmap/; Imputation, https://github.com/CNSGenomics/impute-pipe; HAPI-UR, https://code.google.com/p/hapi-ur/; Impute2, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html; Plink1.9, https://www.cog-genomics.org/plink2; BayesR, https://github.com/syntheke/bayesR; MTG2, https://sites.google.com/site/honglee0707/mtg2.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful comments; the participants of the cohort studies; our colleagues at the Program in Complex Trait Genomics and M. Goddard for comments and suggestions. The authors also wish to thank the staff, contributing research centers and the participants of all studies.

The UK Biobank research was conducted using the UK Biobank Resource under project 12505. The University of Queensland group is supported by the Australian Research Council (Discovery Project 160103860), the Australian National Health and Medical Research Council (1080157, 1078037, 1048853, 1050218, and 1113400), and the NIH (R21ESO25052-01 and PO1GMO99568). J.Y. is supported by a Charles and Sylvia Viertel Senior Medical Research Fellowship. M.R.R. is supported by the University of Lausanne. D.C. and M.J. were supported by the Swedish Research Council (421-2013-1061), the Ragnar Söderberg Foundation (E9/11), and the Jan Wallander and Tom Hedelius Foundation (P2015-0001:1). The ARIC study is carried out as a collaborative study supported by the US National Heart, Lung, and Blood Institute (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The Estonian Genome Centre of University of Tartu Study was supported by EU Horizon 2020 grants 692145, 676550, and 654248; Estonian Research Council Grant IUT20-60, NIASC, EIT–Health; NIH BMI grant 2R01DK075787-06A1; and the European Regional Development Fund (project 2014-2020.4.01.15-0012 GENTRANSMED). The Health and Retirement Study is supported by the US National Institute on Aging (NIA; U01AG009740).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

The genotyping was funded separately by the NIA (RC2 AG036495 and RC4 AG039029) and was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the HRS data were performed by the Genetics Coordinating Center at the University of Washington. The LifeLines Cohort Study and generation and management of GWAS genotype data for the LifeLines Cohort Study were supported by the Netherlands Organization of Scientific Research NWO (175.010.2007.006); the Economic Structure Enhancing Fund of the Dutch government; the Ministry of Economic Affairs; the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the Northern Netherlands Collaboration of Provinces; the Province of Groningen; University Medical Center Groningen; the University of Groningen; the Dutch Kidney Foundation; and the Dutch Diabetes Research Foundation. The Nurses Health Study (NHS) and Health Professionals Follow-up Studies (HPFS) received funding support for the GWAS of Gene and Environment Initiatives in Type 2 Diabetes through the NIH Genes, Environment and Health Initiative (GEI) (U01HG004399). The human subjects participating in the GWAS derived from NHS and HPFS and these studies are supported by National Institutes of Health grants CA87969, CA55075 and DK58845. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies, GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Funding support for genotyping, which was performed at the Broad Institute of MIT and Harvard, was provided by the NIH GEI (U01HG004424). The Swedish Twin Registry (TWINGENE) was supported by the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), NIH DK U01-066134, the Swedish Foundation for Strategic Research (SSF), and the Heart and Lung Foundation (20070481).

AUTHOR CONTRIBUTIONS

M.R.R. and P.M.V. conceived and designed the study. M.R.R. conducted all analysis, with contributions from G.E., G.M., L.R.L.-J., D.C., and M.A.T. G.M. developed the BayesR software, and J.Y. developed the GCTA software. The LifeLines Cohort Study, Z.Z., I.M.N., J.V.v.V.-O., H.S., T.E., L.M., R.M., A.M., P.K.E.M., N.L.P., E.I., M.J., J.Y., and D.C. provided study oversight, sample collection, and management. M.R.R. and P.M.V. wrote the manuscript. All authors reviewed and approved the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Visscher, P.M., Hill, W.G. & Wray, N.R. Heritability in the genomics era—concepts and misconceptions. *Nat. Rev. Genet.* **9**, 255–266 (2008).
2. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, 1998).
3. Vinkhuyzen, A.A.E., Wray, N.R., Yang, J., Goddard, M.E. & Visscher, P.M. Estimation and partition of heritability in human populations using whole-genome analysis methods. *Annu. Rev. Genet.* **47**, 75–95 (2013).
4. Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
5. Wray, N.R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515 (2013).
6. Schousboe, K. *et al.* Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries. *Twin Res.* **6**, 409–421 (2003).
7. Elks, C.E. *et al.* Variability in the heritability of body mass index: a systematic review and meta-regression. *Front. Endocrinol. (Lausanne)* **3**, 29 (2012).
8. Coady, S.A. *et al.* Genetic variability of adult body mass index: a longitudinal assessment in Framingham families. *Obes. Res.* **10**, 675–681 (2002).
9. Vattikuti, S., Guo, J. & Chow, C.C. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet.* **8**, e1002637 (2012).
10. Zaitlen, N. *et al.* Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet.* **9**, e1003520 (2013).
11. Visscher, P.M. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* **2**, e41 (2006).
12. Hemani, G. *et al.* Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. *Am. J. Hum. Genet.* **93**, 865–875 (2013).
13. Polderman, T.J.C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
14. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
15. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).
16. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
17. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
18. Zhu, Z. *et al.* Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am. J. Hum. Genet.* **96**, 377–385 (2015).
19. Magnusson, P.K.E. & Rasmussen, F. Familial resemblance of body mass index and familial risk of high and low body mass index. A study of young men in Sweden. *Int. J. Obes. Relat. Metab. Disord.* **26**, 1225–1231 (2002).
20. Rietveld, C.A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
21. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M. & Wray, N.R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism–derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540–2542 (2012).
22. Kirkpatrick, M., Hill, W.G. & Thompson, R. Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. *Genet. Res.* **64**, 57–69 (1994).
23. Kirkpatrick, M., Lofsvold, D. & Bulmer, M. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* **124**, 979–993 (1990).
24. Meyer, K. Estimating covariance functions for longitudinal data using a random regression model. *Genet. Sel. Evol.* **30**, 221 (1998).
25. Meyer, K. & Kirkpatrick, M. Up hill, down dale: quantitative genetics of curvaceous traits. *Phil. Trans. R. Soc. Lond. B* **360**, 1443–1455 (2005).
26. Robinson, M.R., Pilkington, J.G., Clutton-Brock, T.H., Pemberton, J.M. & Kruuk, L.E.B. Environmental heterogeneity generates fluctuating selection on a secondary sexual trait. *Curr. Biol.* **18**, 751–757 (2008).
27. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
28. Haberstick, B.C. *et al.* Stable genes and changing environments: body mass index across adolescence and young adulthood. *Behav. Genet.* **40**, 495–504 (2010).
29. Stunkard, A.J., Foch, T.T. & Hrubec, Z. A twin study of human obesity. *JAMA* **256**, 51–54 (1986).
30. Silventoinen, K., Kaprio, J. & Lahelma, E. Genetic and environmental contributions to the association between body height and educational attainment: a study of adult Finnish twins. *Behav. Genet.* **30**, 477–485 (2000).
31. Winkler, T.W. *et al.* The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. *PLoS Genet.* **11**, e1005378 (2015).
32. Ng, M. *et al.* Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 766–781 (2014).
33. Choquet, H. & Meyre, D. Genetics of obesity: what have we learned? *Curr. Genomics* **12**, 169–179 (2011).
34. Qi, Q. *et al.* Sugar-sweetened beverages and genetic risk of obesity. *N. Engl. J. Med.* **367**, 1387–1396 (2012).
35. Qi, Q. *et al.* Television watching, leisure time physical activity, and the genetic predisposition in relation to body mass index in women and men. *Circulation* **126**, 1821–1827 (2012).
36. Young, A.I., Wauthier, F. & Donnelly, P. Multiple novel gene-by-environment interactions modify the effect of *FTO* variants on body mass index. *Nat. Commun.* **7**, 12724 (2016).
37. Yang, J. *et al. FTO* genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267–272 (2012).
38. Rosenquist, J.N. *et al.* Cohort of birth modifies the association between *FTO* genotype and BMI. *Proc. Natl. Acad. Sci. USA* **112**, 354–359 (2015).
39. Tyrrell, J. *et al.* Gene–obesogenic environment interactions in the UK Biobank study. *Int. J. Epidemiol.* **46**, 559–575 (2017).
40. Fusi, N., Lippert, C., Lawrence, N.D. & Stegle, O. Warped linear mixed models for the genetic analysis of transformed phenotypes. *Nat. Commun.* **5**, 4890 (2014).
41. Moser, G. *et al.* Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genet.* **11**, e1004969 (2015).
42. Robinson, M.R. *et al.* Genetic evidence of assortative mating in humans. *Nat. Hum. Behav.* **1**, 16 (2017).
43. Ober, C., Loisel, D.A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nat. Rev. Genet.* **9**, 911–922 (2008).
44. Weiss, L.A., Pan, L., Abney, M. & Ober, C. The sex-specific genetic architecture of quantitative traits in humans. *Nat. Genet.* **38**, 218–222 (2006).
45. Randall, J.C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* **9**, e1003500 (2013).
46. Yang, J. *et al.* Genome-wide genetic homogeneity between sexes and populations for human height and body mass index. *Hum. Mol. Genet.* **24**, 7445–7449 (2015).
47. Traglia, M. *et al.* Genetic mechanisms leading to sex differences across common diseases and anthropometric traits. *Genetics* **205**, 979–992 (2017).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

ONLINE METHODS

Estimating heritability of BMI and height from sibling data. To explore the robustness of twin-based heritability estimates for BMI and height to the inclusion of other types of siblings, we used a large sample of Swedish brothers born during 1950–1969. The sample contains seven types of siblings: monozygotic twins (n = 1,154), dizygotic twins (n = 1,601), full siblings reared together (n = 151,699 for BMI, n = 151,789 for height), full siblings reared apart (n = 1,033), half-siblings reared together (n = 4,874 for BMI, n = 4,880 for height), and half-siblings reared apart (n = 11,547 for BMI, n = 11,566 for height). To obtain this sample, we combined data from Swedish administrative records, taking information about BMI and height from Swedish conscription records. The samples have previously been used to investigate the heritability of educational attainment and cognitive skills[20]. Sample construction draws heavily on previously published work[20,48,49]; details can be found in the **Supplementary Note. Supplementary Table 1** reports pairwise correlations of BMI and height for sibling pairs. The phenotypic correlations were calculated after residualizing the original variable on indicator variables for region of residence (25 in total) and year of birth. From these correlations, we assumed a commonly applied model of additive genetic variance, common environmental variance, and individual-level environmental variance (ACE model) for both BMI and height. The common environment component was estimated from the difference between full and half-siblings who were raised together and apart as

$$C = \frac{d_{FS}/\sigma^2_{d_{FS}} + d_{HS}/\sigma^2_{d_{HS}}}{(1/\sigma^2_{d_{FS}}) + (1/\sigma^2_{d_{HS}})},$$

where d_{FS} is the difference in correlation between full siblings raised together and those raised apart; d_{HS} is the difference in correlation between half-siblings raised together and those raised apart; $\sigma^2_{d_{FS}}$ and $\sigma^2_{d_{HS}}$ are the sampling variances of d_{FS} and d_{HS}, respectively. We assumed assortative mating at equilibrium, with a correlation among spousal pairs (ρ_p) of 0.16 for BMI and 0.2 for height, as reported in a previous study[42]. We then used the coefficients presented previously[2] to estimate h^2 of BMI for different relative pairs as h^2 MZ twin pairs = $\rho_{MZ} - C$; h^2 DZ twin pairs = $2(\rho_{DZ} - C) / AM_{FS}$; h^2 full siblings raised together = $2(\rho_{FS} - C) / AM_{FS}$; h^2 full siblings raised apart = $2\rho_{FS} / AM_{FS}$; h^2 half-siblings raised together = $2(\rho_{HS} - C) / AM_{HS}$; h^2 half-siblings raised apart = $2\rho_{HS} / AM_{HS}$. $AM_{FS} = 1 + \rho_p h^2$ and $AM_{HS} = 1 + 2\rho_p h^2 + \rho_p{}^2 h^2$. This provides a full method of moments approach comparable to that gained from a maximum likelihood ACE approach. We then repeated the estimation assuming the absence of assortative mating to show that our conclusions are not driven by assortative mating approximations (**Supplementary Tables 1** and **2** and **Fig. 1**). Standard errors (s.e.) are approximated as $\left(1-\hat{\rho}^2\right)/\sqrt{P}$, with P the number of pairs, and thus for example the s.e. of the heritability estimate for MZ twin pairs is $1-\left(\rho_{FS}-C\right)^2\Big)/\sqrt{P}$.

Large population samples. We utilize large publicly available phenotype–genotype data sets of unrelated individuals in a series of experimental designs that are unbiased of environmental confounding to test for changes in the genetic basis of BMI across ages and environments and to compare the underlying genetic architecture of both traits.

AHTHEL composite cohorts. We accessed data from four cohorts through dbGaP: The Atherosclerosis in Communities (ARIC) study, the Nurses Health Study (NHS), the Health Professionals Follow-up Study (HPFS), and the Health and Retirement Study (HRS). We combined this with data from the TwinGene, the Estonian Genome Center of University of Tartu Study (EGCUT), and the LifeLines studies (**Supplementary Table 4**). A summary description of the sample sizes, genotyping platforms, quality-control criteria for the genotype data, and imputation process has been described previously[15]. We selected 1,233,988 HapMap3 SNPs from the imputed genotype data.

To avoid including close relatives in the sample, we estimated the genetic relatedness for pairs of individuals in the combined data set at the HapMap3 SNPs and removed one of each pair of individuals with estimated genetic relatedness $r > 0.05$. We retained 43,407 unrelated individuals of European descent[15] where age and sex were known (**Supplementary Table 3**). BMI and height were recorded for every individual, and within each cohort, we adjusted both phenotypes for age (factor with levels for each age between 20 and 84) and sex differences. BMI and height phenotypes 5 s.d. away from the mean were not included in the analyses. Both phenotypes were then converted to z-scores with 0 mean and variance 1 within each cohort so that when the data were combined all mean cohort effects were removed.

The UK Biobank Study. The UK Biobank Study is a prospective cohort study of more than 500,000 individuals from across the United Kingdom. Participants, aged between 37 and 73, were invited to one of 22 centers across the United Kingdom between 2006 and 2010. After quality-control (QC) procedures were applied (see URLs), the interim UK Biobank data release contained genotypes for 152,736 samples that passed sample QC (~99.9% of total samples). We used the imputed genotype data provided as part of the data release where imputation was from a merged reference panel from the UK10K and 1000 Genomes data (see URLs), providing a data set with 73,355,667 SNPs, short indels, and large structural variants in 152,249 individuals. Selecting out only SNPs with imputation info score >0.3 and minor allele count ≥5 gave ~40 million SNPs in 152,249 individuals.

Principal component analysis and self-declared ethnicity were used to derive a 'white British' subset of samples. In addition, samples were excluded if they had (i) at least one identified closely related sample ($r > 0.1$); (ii) a genetically inferred gender that did not match the self-reported gender; or (iii) ~500 extreme heterozygosity or missing genotype outliers. These filters resulted in a data set with 112,338 individuals, and further exclusion of one individual from a pair with $r > 0.05$ using GCTA (see URLs) resulted in a final sample of 107,488 individuals with measures of BMI and height. From this sample, we then selected out 1,162,900 HapMap3 SNPs. We selected only the first recorded measures of BMI and height and adjusted both phenotypes for age (factor with levels for each age between 40 and 73) and sex differences. BMI and height phenotypes 5 s.d. away from the mean were not included in the analyses. Both phenotypes were then converted to z-scores with 0 mean and variance of 1.

Estimating genetic variance and covariance across a gradient. *MV-GREML.* Consider a population study where a phenotype, y, is measured on different individuals at different points along a continuous gradient (age, lifestyle factor, or a measure of the environment):

$$\begin{aligned}\mathbf{y}_1 &= \mathbf{X}_1\mathbf{b}_1 + \mathbf{Z}_1\mathbf{g}_1 + \mathbf{e}_1\\ \mathbf{y}_2 &= \mathbf{X}_2\mathbf{b}_2 + \mathbf{Z}_2\mathbf{g}_2 + \mathbf{e}_2\\ &\vdots\\ \mathbf{y}_t &= \mathbf{X}_t\mathbf{b}_t + \mathbf{Z}_t\mathbf{g}_t + \mathbf{e}_t\end{aligned} \quad (1)$$

where $\mathbf{y}_i$ is a vector of phenotypes measured at point i (i = 1, …, t) along a continuous gradient, $\mathbf{b}_i$ is a vector of fixed effects, $\mathbf{g}_i$ is a vector of additive genetic values for individuals, and $\mathbf{e}_i$ are the residuals. The random effects, $\mathbf{g}_i$ and $\mathbf{e}_i$, are assumed to be normally distributed with mean zero, and $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices relating the individual measurements to the fixed and random effects respectively, with $\mathbf{Z}_i$ of dimension N_i x N, containing 0 columns for individuals not measured for a given phenotype, with N the number of individuals.

The variance–covariance matrix of all the observations can be written as

$$\mathbf{V} = \mathrm{var}\left(\mathbf{y}_{\boldsymbol{i}}\right) = \begin{bmatrix} \mathbf{Z}_1\mathbf{A}\sigma^2_{g_1}\mathbf{Z}_1' + \mathbf{I}\sigma^2_{e_1} & \cdots & \mathbf{Z}_1\mathbf{A}\sigma_{g_{1,t}}\mathbf{Z}_t' \\ \vdots & \ddots & \vdots \\ \mathbf{Z}_1\mathbf{A}\sigma_{g_{1,t}}\mathbf{Z}_t' & \cdots & \mathbf{Z}_t\mathbf{A}\sigma^2_{g_t}\mathbf{Z}_t' + \mathbf{I}\sigma^2_{e_t} \end{bmatrix} \quad (2)$$

where $\mathbf{A}$ is a genomic relationship matrix calculated from the SNP markers with l,k^{th} elements:

$$\mathbf{A}_{l,k} = \frac{1}{N}\sum_{j=1}^{M}\frac{\left(x_{j,l}-2p_j\right)\left(x_{j,k}-2p_j\right)}{2p_j(1-p_j)},$$

where x_j is a SNP marker dosage describing the number of minor alleles (0,1,2) of SNP j, p_j is the minor allele frequency of SNP j, M is the number of markers and $\mathbf{I}$ is an identity matrix of dimension $N \times N$. The parameters $\sigma^2_{g_t}$ and $\sigma^2_{e_t}$ denote the additive genetic variance tagged by SNP markers and residual

doi:10.1038/ng.3912

variance respectively and $\sigma_{g_{i,t}}$ denotes the additive genetic covariance tagged by SNP markers among phenotypic measurements at different points along the continuous gradient. For the data we apply this model to, individuals only have one measurement at a single time point, so there is no residual covariance, however, $\sigma_{g_{i,t}}$ is estimated because individuals across the points of measurement along the continuous gradient vary in their similarity at SNP markers, which provides a direct estimate of the genome-wide correlation of SNP effects across a continuous gradient. This results in an estimated genetic variance-covariance matrix **G**:

$$\mathbf{G}=\begin{bmatrix} \sigma^2_{g_1} & \cdots & \sigma_{g_{1,t}} \\ \vdots & \ddots & \vdots \\ \sigma_{g_{1,t}} & \cdots & \sigma^2_{g_t} \end{bmatrix} \quad (3)$$

As described[50], the log likelihood of equation (1) is

$$\ln L=-\frac{1}{2}\left[\ln|\mathbf{V}|+\ln\left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right|+\mathbf{y}'\mathbf{P}\mathbf{y}\right]$$

with

$$\mathbf{X}=\begin{bmatrix} \mathbf{X}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{X}_t \end{bmatrix}$$

and

$$\mathbf{y}=\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_t \end{bmatrix},$$

and equation (1) can be estimated using residual maximum likelihood and a direct average information algorithm (AIREML)[50].

RR-GREML. Testing for genome-wide interaction effects across a gradient requires testing for differences among the elements of **G**. This can be done in a single LRT framework through the use of a covariance function approach[22–26] where **G** is modeled with a reduced number of parameters using the model

$$\mathbf{y}=\mathbf{X}\mathbf{b}+\mathbf{Z}\alpha+\mathbf{e} \quad (4)$$

where **y** is a vector of observations measured on individual, **b** is a vector of fixed effects, α is a vector of additive genetic random regression coefficients, and **e** is the residual error. **Z** is a design matrix containing the elements Φ pertaining to the t measurement points in the data. The matrix Φ of order t x k contains orthogonal polynomial coefficients evaluated at t standardized measurement points and is calculated as $\Phi = \mathbf{M}\Lambda'$ where Λ is the matrix of polynomials of order $k \times k$ and **M** is the standardized covariate to 0 mean and values between −1 and 1. The continuous covariate is standardized as:

$$m=-1+2\left(\frac{t-\min(t)}{\max(t)-\min(t)}\right).$$

The corresponding variances are $\mathrm{var}(\alpha)=\mathbf{K}\otimes\mathbf{A}$, where **K** of order k contains the variance–covariance component α and is estimated directly from the data within the random regression model[22–26] of equation (4), which has been used in animal and plant breeding and evolutionary biology when testing repeated measures data containing close relatives. The residual variances are estimated as a diagonal matrix **R** of dimension $t \times t$, which is appropriate, as independent individuals are used across the t points. As previously described in ref. 50, an AIREML algorithm is used to estimate **K** and **R**. A genetic covariance function of order k, where $k < t$ can then be estimated as:

$$\hat{\mathbf{G}}=\Phi\mathbf{K}\Phi' \quad (5)$$

For example, for a covariate with five measurement points and $k = 1$, $\mathbf{K}=\left[\sigma^2_{g_I}\right]$, $\mathbf{\Lambda}=[0.7071]$ and $\mathbf{M}=[1,1,1,1,1]'$, where $\sigma^2_{g_I}$ is a single parameter representing a constant estimate of σ^2_g across the five measurement points. In this example if $k = 2$ then

$$\mathbf{K}=\begin{bmatrix} \sigma^2_{g_I} & \sigma_{g_I g_S} \\ \sigma_{g_I g_S} & \sigma^2_{g_S} \end{bmatrix},\quad \mathbf{\Lambda}=\begin{bmatrix} 0.7071 & 0 \\ 0 & 1.2247 \end{bmatrix}$$

and

$$\mathbf{M}=\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & -0.5 & 0 & 0.5 & 1 \end{bmatrix}',$$

where $\sigma^2_{g_I}$ is an intercept parameter, $\sigma^2_{g_S}$ is a slope parameter representing the change in σ^2_g across the five measurement points, and $\sigma_{g_I g_S}$ is the covariance of intercept and slope terms.

The model of equation (4) reduces a t dimensional problem of equation (1) into a k dimensional problem, with the resulting random regression coefficients being transformed into estimates of $\hat{\mathbf{G}}$ in equation (5). A model of $k = t - 1$ is equivalent to the full $t \times t$ multivariate MV-GREML model but with a reduced number of parameters. Here we always first fit the full MV-GREML model of equation (1) to estimate **G** and then reanalyze the data using the RR-GREML model of equation (4) to test for changing patterns of genetic variance-covariance across t points. To present the results, we use the estimates in $\hat{\mathbf{K}}$ to calculate $\hat{\mathbf{G}}$ and we calculate the approximate s.e.m. using $\mathrm{var}\left(\hat{\mathbf{K}}\right)$, which are given in the inverse of the AI matrix[51]. Equations (1) and (4) were fit to the data using the software program MTG2 (ref. 50; see URLs).

We conduct model testing by comparing the model fit to the data of equation (4) with a zero-order polynomial ($k = 0$) to equation (4) with a first-order polynomial ($k = 1$). To do this we use a

$$LRT=-2\left(logLK_{\mathrm{model1}}-logLK_{\mathrm{model2}}\right),$$

where *log*LK is the model log likelihood. The value obtained from a LRT approximates a chi-square distribution with degrees of freedom equal to the number of additional parameters that are estimated in model 2 as compared to model 1, and from this we obtain a P value. A zero-order polynomial RR-GREML model estimates a single genetic parameter, and a first-order polynomial RR-GREML model estimates three parameters (intercept, slope, and covariance of the slope and intercept); thus, we use 2 df in our LRT, which is the most conservative approach[52]. This testing approach determines whether there is significant evidence of a change in genetic (co)variance across measurement points, which tests whether there is evidence of genotype–covariate interaction across the entire range of measurement points. We also provide Akaike information criterion ($\mathrm{AIC} = -2logLK + 2p$) with p the number of parameters estimated from the model and smaller values indicating improved fit to the data.

GCI-GREML. Third, we also reanalyze the data to estimate the proportion of phenotypic variance contributed by genotype–covariate interaction effects, using a model proposed by Yang *et al.*[27]. This approach estimates the coheritability of the phenotype across t measurement points, and estimates the proportion of variance attributable to genotype–covariate interaction effects (GCI-GREML model) simultaneously using the model

$$\mathbf{y}=\mathbf{X}\mathbf{b}+\mathbf{Z}\mathbf{g}+\mathbf{Z}\mathbf{g}_{gci}+\mathbf{e} \quad (6)$$

where $\mathbf{g}_{gci}$ is a vector of genotype–covariate interaction effects for individuals and

$$\mathbf{V}=\mathbf{Z}\mathbf{A}_g\sigma^2_g\mathbf{Z}'+\mathbf{Z}\mathbf{A}_{gci}\sigma^2_{gci}\mathbf{Z}'+\mathbf{I}\sigma^2_e,$$

with $\mathbf{A}_g=\mathbf{A}_{gci}$ for individuals in the same measurement point and $\mathbf{A}_{gci}=0$ for individuals of different measurement points. This approach provides a test for the presence of gene–covariate interaction by comparing the fit to the data of a GCI-GREML model (fitting a genetic variance component and a genetic interaction variance component) and a null GREML model (fitting a single genetic variance component) by LRT. We show through theory that a GCI-GREML approach provides an appropriate testing framework for gene–covariate interactions and estimates the variance contributed by genotype–covariate interaction effects (**Supplementary Note**).

We have considered a covariate to be discrete or continuous but with sampling points along the distribution. If that is not the case and a covariate is truly continuous in nature with individuals represented across all values, then it is required that the data be 'blocked' into groups to create a series of points at which sufficient sample size is available to provide reasonable estimates of the variance parameters.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

Genotype–age interaction effects for BMI. We divided the AHTHEL sample into five age groups with the aim of having groups representative of different stages of adult life (**Supplementary Table 4**). We first estimate the full multivariate REML model of equation (1) for both BMI and height, and we then reanalyzed the same data using equation (4), with five residual variance components and $k = 1$ to test for any change in additive genetic (co)variance across ages. We then estimated equation (6) both across the entire data set and then between each pair of age groups. This gives a total of 26 LRT (2 MV-GREML versus GREML tests, 2 RR-GREML $k = 1$ versus RR-GREML $k = 0$ tests, 2 GCI-GREML versus GREML tests for the entire data set, and 22 GCI-GREML versus GREML tests for each pair of age groups). In this sample, individuals were only measured once, and thus the residual covariance was set to 0 in all models. In all models, we included the first 15 principal components of the HapMap3 SNP data as fixed effects to control for population stratification. To correct for mean–variance relationships, and to adjust for heterogeneity in phenotypic variance before testing for the presence of gene–covariate interactions, we standardized BMI measures within each age group with a rank inverse normal transformation to remove differences in phenotypic mean and variance across age. This means that we estimate genotype–age interaction effects on the heritability scale and that we will detect genotype–age interaction only if the genetic effects change in a way that does not scale proportionally with any changes in phenotypic variance across ages. To test heterogeneity in variances across age for both BMI and height, we then repeat the analyses without rank inverse normal transformation within each age group, with a further 4 LRTs (2 RR-GREML $k = 1$ versus RR-GREML $k = 0$ tests, 2 GCI-GREML versus GREML tests for the entire data set).

We then repeated the analysis of BMI in the UK Biobank sample, which contained individuals measured between the ages of 46 and 73, dividing individuals into deciles of the age distribution to create 10 groups of approximately equal sample size. The age ranges were: (i) 9,655 individuals aged 40 to 44; (ii) 10,379 individuals aged 45 to 48; (iii) 9,205 individuals aged 49 to 51; (iv) 13,669 individuals aged 52 to 55; (v) 7,689 individuals aged 56 and 57; (vi) 8,590 individuals aged 58 and 59; (vii) 11,211 individuals aged 60 and 61; (viii) 11,019 individuals aged 62 and 63; (ix) 14,559 individuals aged 64 to 66; and (x) 11,612 individuals aged 67 and above. We included the first 15 principal components supplied by the UK Biobank as fixed effects in each model to control for population stratification. To correct for mean–variance relationships and to adjust for heterogeneity in phenotypic variance before testing for the presence of gene–covariate interactions, we standardized BMI measures within each age group with a rank inverse normal transformation to remove differences in phenotypic mean and variance across age. Again, this means that we estimate genotype–age interaction effects on the heritability scale and that we will detect genotype–age interaction only if the genetic effects change in a way that does not scale proportionally with any changes in phenotypic variance across ages. We first estimate the full multivariate REML model of equation (1), and we then reanalyzed the same data using equation (4), with ten residual variance components and $k = 1$ to test for any change in additive genetic (co)variance across ages. Finally, we then also estimated equation (6) across the whole data set and among the pairs of age groups. This gives a total of 48 LRT.

Genotype–environment interaction effects for BMI. We identified a series of self-reported lifestyle factors that influence BMI variation in the UK Biobank data, using a linear model (**Supplementary Table 5**). Prior to analysis we adjusted BMI for sex and age and standardized the values to a z-score. This resulted in a set of 97,510 individuals who had full phenotypic records at all 11 lifestyle factors that significantly influenced BMI. To ensure that all mean effects were removed, we then adjusted BMI by each of the 11 lifestyle factors by taking the standardized residuals from the linear model.

We then used the lifestyle factors identified as significantly influencing BMI and estimated the correlation among measures (**Supplementary Fig. 5**). On the basis of correlations, we combined alcohol consumption variables of weekly frequency of alcohol intake with weekly number of drinks consumed into a single alcohol consumption variable, grouped the self-report exercise variables of moderate and vigorous weekly exercise into a single exercise variables, and combined binary smoking variable and number of pack years as a proportion of life span to give a single pack year variable. This gave a total of eight lifestyle variables; a full description of the construction of these variables and their groupings is given in the **Supplementary Note**. We then estimated equation (1), equation (4), and equation (6) for each of the eight variables, giving a total of 163 LRT. We included the first 15 principal components supplied by the UK Biobank as fixed effects in each model to control for population stratification. For each of the eight lifestyle variables, we correct for mean–variance relationships and adjust for heterogeneity in phenotypic variance before testing for the presence of gene–covariate interactions, by standardizing the residual BMI measures within each group with a rank inverse normal transformation. Again, this means that we estimate genotype–environment interaction effects on the heritability scale and that we will only detect interaction effects if the genetic effects change in a way that does not scale proportionally with any changes in mean or in phenotypic variance across lifestyle factors. To test heterogeneity in variances across environments we then repeat the MV-GREML and RR-GREML analyses without standardizing the phenotype within each group, using eight LRT for the RR-GREML analyses.

Comparing the genetic basis of BMI and height in a Bayesian mixture model. To model the genetic architecture of BMI and height, we used a Bayesian mixture model that fits all genomic markers simultaneously. In GWAS, genomic marker effects are analyzed one at a time, which fails to account for the effects of other SNPs, resulting in increased error variance and decreased power to detect true associations. Additionally, because SNPs are treated as fixed effects and multiple testing occurs, stringent association thresholds result in many false negatives and over-estimated effect sizes for SNPs declared genome-wide significant. Together, this results in only a fraction of the heritability explained by genome-wide significant SNPs and low predictive power in personalized medicine.

An alternative approach is estimating the effects of all genomic markers on a phenotype together, which requires variable selection, either by discarding unimportant predictors, or by shrinking their effects toward 0, because there are typically more genomic markers than there are individuals within a sample. Here we use the Bayesian mixture model, BayesR[41,53] to dissect genetic variation for BMI and height and to shed light on the genetic architecture underlying these complex traits. All SNPs are fit in the model simultaneously, with effects drawn from a prior distribution that matches the true distribution of SNP effects[54]. We do not know the true distribution of effects, but a mixture of normals can approximate a wide variety of distributions[41,55]. Fitting all SNPs together using mixture distributions is expected to give greater power to detect associations, find fewer false negatives, give unbiased estimates of the larger SNP effects, and provide information about the genetic architecture of the trait from the hyper-parameters of the distribution of SNP effects[41,55]. We *a priori* assumed a mixture of five zero-mean normal distributions of SNP effects (β), where the relative variance for each mixture component is fixed as

$$p\left(\beta_j \mid \pi, \sigma_g^2\right) = \pi_1 \times N(0,0) + \pi_2 \times N\left(0, 10^{-5} \times \sigma_g^2\right) + \pi_3 \times N\left(0, 10^{-4} \times \sigma_g^2\right) + \pi_4 \times N\left(0, 10^{-3} \times \sigma_g^2\right) + \pi_5 \times N\left(0, 10^{-2} \times \sigma_g^2\right) \quad (7)$$

with π the mixture proportions which are constrained to unity and σ_g^2 the additive genetic variance captured by the SNPs. Sparseness is induced by the zero effect and variance of the first mixture. σ_g^2 is estimated as a hyper-parameter from the data.

We estimate this model in the UK Biobank data. To remove redundant marker information from the data[41], we estimated LD among HapMap3 SNP markers and selected SNPs with LD $R^2 < 0.7$. This gave 434,491 SNPs and 107,488 individuals. We did not conduct specific hypothesis testing for each SNP; rather, we assessed the proportion of variance tagged by SNPs that is attributable to each mixture component.

Estimating dominance variance for BMI in the UK Biobank. To model the amount of phenotypic variance captured by dominance effects we used the model of Zhu *et al.*[18], where a univariate GCTA-GREML approach[14] is taken to simultaneously estimate the proportion of variation in BMI variance that can be attributable to additive effects at common HapMap3 SNPs, h^2_{SNP} and the proportion of variation that can be attributable to dominance effects at common HapMap3 SNPs, d^2_{SNP}.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

doi:10.1038/ng.3912

Data availability. Data from the EGCUT and LifeLines studies can be obtained from the authors on request. We also used data from the database of Genotypes and Phenotypes (dbGaP), available under accession codes phs000090 (ARIC study), phs000091 (NHS and HPFS), and phs000428 (HRS).

48. Bjorklund, A., Jantti, M. & Solon, G. in *Unequal Chances: Family Background and Economic Success* (eds. Bowles, S., Gintis, H. & Osborne-Grave, M.) 145–164 (Princeton University Press, 2005).
49. Magnusson, P.K.E. *et al.* The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin Res. Hum. Genet.* **16**, 317–329 (2013).
50. Lee, S.H. & van der Werf, J.H.J. MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics* **32**, 1420–1422 (2016).
51. Fischer, T.M., Gilmour, A.R. & van der Werf, J.H.J. Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. *Genet. Sel. Evol.* **36**, 363–369 (2004).
52. Stram, D.O. & Lee, J.W. Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–1177 (1994).
53. Erbe, M. *et al.* Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* **95**, 4114–4129 (2012).
54. Goddard, M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257 (2009).
55. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* **9**, e1003264 (2013).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.