Journal of the American Heart Association

ORIGINAL RESEARCH

Whole-Genome Approach Discovers Novel Genetic and Nongenetic Variance Components Modulated by Lifestyle for Cardiovascular Health

Xuan Zhou, PhD; Julius van der Werf, PhD; Kristin Carson-Chahhoud, PhD; Guiyan Ni, PhD; John McGrath, MD, PhD; Elina Hyppönen, PhD; S. Hong Lee, PhD*

BACKGROUND: Both genetic and nongenetic factors can predispose individuals to cardiovascular risk. Finding ways to alter these predispositions is important for cardiovascular disease prevention.

METHODS AND RESULTS: We used a novel whole-genome approach to estimate the genetic and nongenetic effects on—and hence their predispositions to—cardiovascular risk and determined whether they vary with respect to lifestyle factors such as physical activity, smoking, alcohol consumption, and dietary intake. We performed analyses on the ARIC (Atherosclerosis Risk in Communities) Study (N=6896–7180) and validated findings using the UKBB (UK Biobank, N=14 076–34 538). Lifestyle modulation was evident for many cardiovascular traits such as body mass index and resting heart rate. For example, alcohol consumption modulated both genetic and nongenetic effects on body mass index, whereas smoking modulated nongenetic effects on heart rate, pulse pressure, and white blood cell count. We also stratified individuals according to estimated genetic and nongenetic effects that are modulated by lifestyle factors and showed distinct phenotype–lifestyle relationships across the stratified groups. Finally, we showed that neglecting lifestyle modulations of cardiovascular traits would on average reduce single nucleotide polymorphism heritability estimates of these traits by a small yet significant amount, primarily owing to the overestimation of residual variance.

CONCLUSIONS: Lifestyle changes are relevant to cardiovascular disease prevention. Individual differences in the genetic and nongenetic effects that are modulated by lifestyle factors, as shown by the stratified group analyses, implies a need for personalized lifestyle interventions. In addition, single nucleotide polymorphism–based heritability of cardiovascular traits without accounting for lifestyle modulations could be underestimated.

Key Words: cardiovascular traits ■ genotype–lifestyle interaction ■ lifestyle ■ residual–lifestyle interaction ■ whole-genome approach

Cardiovascular diseases (CVDs) are the world's number 1 cause of mortality, claiming an estimated total of 17.9 million lives globally in the year 2016 alone—that is 31% of the total deaths in a single year.[1] Managing CVD risk is therefore a top public health priority worldwide. It is estimated that between 20% and 60% phenotypic variability in CVD-related traits such as blood pressure and blood-clotting factors are attributed to additive genetic variation (see ref.[2–6]), and the remaining 40% to 80%, commonly referred to as residual variation, could arise from random measurement errors and systematic nongenetic variation in the epigenome, transcriptome, metabolome, proteome, and microbiome, which are involved in or interact with the translation of genotype to phenotype.

Correspondence to: S. Hong Lee, PhD, Australian Centre for Precision Health, Level 8 SAHMRI Building, City West campus, University of South Australia, Adelaide, South Australia, Australia. 5000. E-mail: hong.lee@unisa.edu.au

Supplementary Materials for this article are available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.119.015661

*Preprint posted on BioRxiv July 14, 2019. https://doi.org/10.1101/700617.

For Sources of Funding and Disclosures, see page 15.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- A novel whole-genome approach reveals that lifestyle factors can modulate genetic and nongenetic effects on cardiovascular traits.

What Are the Clinical Implications?

- Lifestyle changes are relevant to cardiovascular disease prevention, but the associated cardiovascular health benefits are greater for some individuals than others, implying a need for personalized lifestyle interventions.

Nonstandard Abbreviations and Acronyms

ARIC	Atherosclerosis Risk in Communities
G-C	genotype–covariate
GREML	genomic restricted maximum likelihood
MRNM	multivariate reaction norm model
R-C	residual–covariate
RNM	reaction norm model
UKBB	UK Biobank

Given the substantial genetic and nongenetic contributions to CVD risk, identifying ways that modify their effects can have important implications for CVD prevention. In fact, the idea of genotype–environment or genotype–covariate (G-C) interaction is well established for traits such as body mass index (BMI).[7–9] That is, genetic effects vary depending on environmental exposure, such as modifiable lifestyle covariates including smoking, alcohol intake, and physical activity. Similar to G-C interaction to genetic variance, we recently demonstrated that some nongenetic variance components can exist that change with respect to lifestyle covariates, which we termed residual–covariate (R-C) interaction,[10] that is, phenotypic variation with respect to lifestyle covariates that is independent of genetic effects.

Understanding G-C and R-C interactions in the context of cardiovascular traits will not only translate into empowering public messages but also enable personalized lifestyle changes for CVD prevention according to individuals' genetic and nongenetic information, as opposed to a one-fits-all approach that neglects individual differences. Aside from its practical implications, studying G-C and R-C interactions is also of theoretical value as it may offer some insight into missing heritability.[11,12]

To date, G-C interaction estimates for cardiovascular traits are based on a limited number of genetic variants[13–20]; therefore they are likely underestimated. R-C interaction has been largely neglected, leading to potential confounding between G-C and R-C interactions in the presence of genuine R-C interaction.[10] In this article, using a novel whole-genome approach,[10] we extend the current understanding of G-C and R-C interactions on cardiovascular health. Instead of focusing on a few genetic variants with large phenotypic effects, our approach uses all common single nucleotide polymorphisms (SNPs) capturing variation across the entire genome, thereby providing genome-wide estimates of G-C interaction. Furthermore, our approach allows residual variance to vary with respect to a chosen covariate, thereby providing estimates of R-C interaction. By examining G-C and R-C interactions, we identify lifestyle factors that modulate genetic and/or nongenetic effects on traits that are indicative of CVD risk.

METHODS

Simulated data used in this article can be obtained from the authors upon request. Our access to the ARIC (Atherosclerosis Risk in Communities) Study data was under the code phs000090, and access to the UKBB (UK Biobank) data was approved by the UKBB research ethics committee under the reference number 14 575.

Our analyses were based on the following 2 data sets: the ARIC Study and the UKBB. Because of the sensitive nature of the data collected for this study, requests to access the data sets from qualified researchers may be sent to the database of Genotypes and Phenotypes (dbgap-help@ncbi.nlm.nih.gov) and the UKBB (access@ukbiobank.ac.uk). The ARIC Study was chosen for our primary analyses because it covers a wider range of cardiovascular traits than the latter data set. The UKBB, which has a larger sample size than the ARIC Study, was chosen as the validation set. The sample sizes for our analyses depended on the availability of phenotype and genotype data, which varied between studies and cross traits. For the ARIC Study, the sample sizes were between 6896 and 7180. For the UKBB, the sample sizes were between 14 076 and 34 538.

Ethics Statement

The current study was approved by the University of South Australia Human Research Ethics Committee. The ARIC Study was approved by the institutional review boards of all participating institutions, including the University of Minnesota, Johns Hopkins University, University of North Carolina, University of Mississippi Medical Centre, and Wake Forest University. The UKBB was approved by the North West Multi-centre

Research Ethics Committee (11/NW/0382). All ARIC Study and UKBB participants gave written informed consent.

ARIC Study

The ARIC Study is a prospective study on the cause of atherosclerosis, with data collected from up to 5 visits over 15 years from participants of 4 U.S. communities (Forsyth County, NC; Jackson, MS; suburbs of Minneapolis, MN; and Washington County, MD) who were aged between 45 and 64 years in 1987 to 1989.[21] To maximize the sample size for our analyses, we only used data from visit 1 that occurred from 1987 to 1989.

Cardiovascular Traits

A total of 23 cardiovascular health-related traits were selected. Coagulation factors were determined in the ARIC Central Hemostasis Laboratory using previously published procedures.[22] Plasma concentration of fibrinogen was measured by the thrombin–titration method, factor VII and factor VIII activity by clotting assays, and Von Willebrand's factor antigen with an ELISA technique.[23,24] P-R interval, Q-T interval, QRS interval, and Cornell voltage were derived from standard 12-lead electrocardiography.[25,26] Sitting blood pressure (systolic and diastolic) was measured 3 times from the right arm and calculated using the average of the last 2 readings. Pulse pressure was computed as the difference between systolic and diastolic blood pressure.

Lifestyle Covariates

A total of 22 lifestyle covariates were selected. Smoking was indexed by "cigarette years of smoking," derived by multiplying the average number of cigarettes per day with the number of years smoked. Alcohol intake was indexed by usual ethanol intake from wine, beer, and hard liquor in grams per week. Dietary composition was assessed using a 66-item food-frequency questionnaire based on the Willett 61-item questionnaire.[27] The summary measures derived included dietary lipid content, as indexed by the keys score[28,29]; daily dietary intake of fiber; monounsaturated, polyunsaturated, and saturated fatty acids; total fat carbohydrate, protein, potassium, calcium, and magnesium; total daily energy intake; and percentages of daily total energy intake from monounsaturated, polyunsaturated, and saturated fatty acids, total fat, carbohydrate, and protein. Physical activity was assessed in work, sports, and leisure domains using a modified Baecke questionnaire.[30,31] Only the summary scores from the sports and leisure questions were used. The score for sports is a summary of the following: (1) the frequency, duration, and an assigned intensity of the sports reported by participants and (2) 3 additional questions on frequency of sweating, general frequency of playing sports, and a self-rating of the amount of leisure time physical activity compared with others of the same age.[32] The score for leisure is a summary of the frequency of watching television (scored inversely), walking, bicycling, and walking/biking to work or shopping.[32]

Genotyping Data

The ARIC Study genotype data set contains 609 441 SNPs that are genotyped for 8291 participants. We first selected autosomes from white European participants then applied standard quality control procedures to the selected data set. This involved (1) excluding SNPs with a genotyping rate less than 95%, ones that failed the Hardy–Weinberg test at the 0.0001 level or had a frequency less than 0.01; (2) excluding individuals who were missing 5% of genotype data; and (3) removing related individuals by excluding 1 person at random using a Bernoulli distribution with a selection probability of 0.5, from each pair that had an estimated genomic relationship[33] greater than 0.05. Eventually, 586 257 SNPs and 7513 individuals remained. Among these individuals, 6896 to 7180 have non missing phenotypic records to be used in the analyses of the 23 traits.

UK Biobank

The UKBB contains health-related data from ≈500 000 participants aged between 40 and 69 years who were recruited throughout the United Kingdom between 2006 and 2010.[34] For validation purposes, we only selected cardiovascular-related phenotypes and lifestyle covariates that overlapped with the ARIC Study data set, which included BMI, waist-to-hip ratio, heart rate, white blood cell count, diastolic and systolic blood pressure, pulse pressure, high-density lipoprotein (HDL) cholesterol level, apolipoprotein a1 level, smoking (pack years of smoking as proportion of life span exposed to smoking), alcohol intake (average weekly intake of all types), physical activity (metabolic equivalent minutes for walking, moderate activity, vigorous activity, and all types[35]), and dietary composition (polyunsaturated fatty acid, saturated fatty acid, and total energy intake).

Genotyping Data

The second release of the UKBB genotyping data set was used. Before quality control, the dataset contains 92 693 895 imputed autosomal SNPs of which genotypes are available for 488 377 individuals. We selected the third phase of the International HapMap project (HapMap3) SNPs from individuals

of white British ancestry only and applied the same quality control procedures as for the ARIC Study genotyping data set. Only HapMap3 SNPs were selected because they were shown to yield reliable and robust estimates of SNP-based heritability and genetic correlation.[36–38] In addition, ambiguous and duplicated SNPs and SNPs with an information score (used to index the quality of genotype imputation) <0.6 were excluded. We computed the genomic relationship matrix of all observations and excluded population outliers, defined as individuals who have a score outside 3 SD on either the first or second principal component of the genomic relationship matrix. From the remaining participants, only those who were part of the first release of the UKBB genotyping data (≈150 000 individuals) were selected for the purpose of reducing computational burden. This subset of participants therefore has 2 versions of imputed genotyping records, 1 from each release, which enabled the computation of discordance rates between the 2 versions for each SNP across individuals and for each individual across SNPs. SNPs and individuals who have a discordance rate >0.05 were excluded. Eventually, 1 130 918 SNPs for 66 281 participants remained. Among these participants, only 14 076 to 34 538 have phenotype data available for analysis.

Statistical Analysis

Interaction Effects Detection

To estimate the variances of G-C and R-C interaction effects, we used multivariate reaction norm models (MRNMs),[10] where the 2 types of interaction effects are treated as random. Details of MRNMs can be found in Data S1. Briefly, under this approach, the presence of a type of interaction effect is evidenced by its nonzero variance, and we declared nonzero variances of interaction effects when the full model, that is, a MRNM that assumes the presence of G-C and R-C interactions, had a better fit than the null model, that is, a MRNM that assumes no G-C and R-C interactions. The full model is illustrated graphically in Figure 1, simulation studies used to calibrate our model comparison approach are documented in Data S2 (Figures S1–S4, Tables S1 and S2), and the justification for the model comparison approach is included in Data S3 (Figure S5, Tables S3 and S4). All model comparisons were based on likelihood ratio tests. For our primary analyses, we applied the full model versus the null model comparison approach to the ARIC Study data set. To validate significant results that emerged from the ARIC Study data set, we repeated the analyses using the UKBB for cardiovascular traits where the 2 data sets overlap.

Of note, the full model versus the null model comparison method on its own does not disentangle orthogonal G-C or R-C interaction effects. As documented in Data S3 (Figure S5, Tables S3 and S4), we initially considered a series of model comparisons to disentangle orthogonal G-C or R-C interaction from overall interactions, but the power of this approach was low (< 11% for interaction effects of a small size; Data S3, Figure S5, Tables S3 and S4). Subsequently, we used the full model comparison versus the null model comparison to detect overall interactions and used parameter estimates from the full model to quantify G-C and R-C interactions. This strategy is based on 2 observations from our simulation studies (Data S2, Figures S1–S4, Tables S1 and S2). First, the full model versus the null model comparison in general had a high power of detecting any interaction type or both (>84% even for small interaction effects; Data S2, Figures S1–S4, Tables S1 and S2). Second, the full model consistently yielded unbiased variance component estimates under all simulation scenarios and parameter settings (Data S2, Figures S1–S4, Tables S1 and S2). Importantly, an interaction effect on its own—whether it is a G-C interaction or an R-C interaction—has important clinical relevance to cardiovascular risk reduction (as shown later).

Phenotype Adjustment

Before fitting MRNMs, all selected cardiovascular traits were adjusted for demographic and lifestyle variables using linear models that regressed phenotypes on demographic and lifestyle variables. The demographic variables included age, sex, education level, marital status, field center identification, and population structure, as measured using the first 15 principal components of the estimated genomic relationship matrix. All lifestyle variables described in the previous section were included for the adjustment. Depending on the cardiovascular trait, some additional adjustment factors were also included. For resting heart rate, blood pressure measures, electrocardiography variables, and coagulation factors, additional adjustment factors included hypertension, defined as systolic blood pressure ≥140 or diastolic blood pressure ≥90, and hypertension-lowering medication use. For total cholesterol and triglycerides levels, additional adjustment factors were hypertension, hypertension-lowering medication use, cholesterol-lowering medication within 2 weeks, and medications that secondarily affect cholesterol. As the second trait in the multivariate reaction normal model (see the second part of Equation 1 in Data S1 where lifestyle covariate is on the left side of the equation), lifestyle covariates were also adjusted in the same way as the cardiovascular trait in the first part of Equation 1 in Data S1.

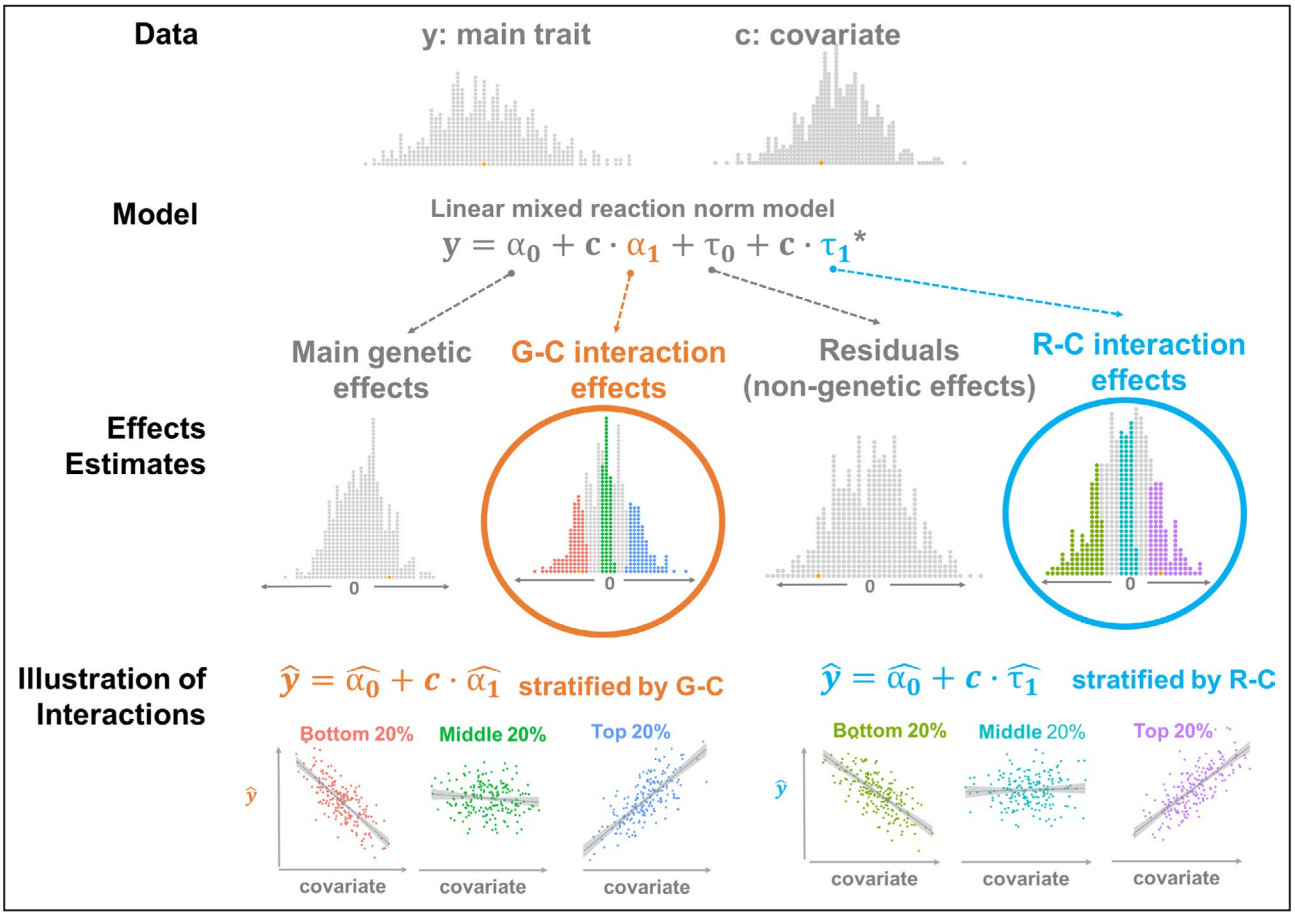


Figure 1. A schematic illustrating the working of the linear mixed reaction norm model used to study genotype–covariate (G-C) and residual–covariate (R-C) interaction effects.

Given phenotypic data of a main trait and a covariate, a reaction norm model that assumes G-C and R-C interactions (ie, a full model) decomposes the phenotypic variance of the main trait into genetic variance, residual variance, and their subcomponents that are modulated by the covariate, that is, variance of G-C interaction effects and variance of R-C interaction effects. The model then generates per-individual estimates of main genetic effects, G-C interaction effects, residuals, and R-C interaction effects, which can be used to compute estimated phenotypes of the main trait, denoted as ŷ. When stratified according to interaction effect estimates, ŷ and covariate shows different relationships across stratified groups, illustrating G-C and R-C interaction effects. For simplicity, only the key part of the linear mixed reaction norm model is shown. *Note that α_1 and τ_1 are n×1 vectors of interaction effects of 2 types, that is, G-C and R-C interactions, respectively, which serve to capture the heterogeneity of genetic variance and the heterogeneity of residual variance across different values of a given covariate. The 2 effects have different variance–covariance structures. Specifically, $\text{var}(\alpha_1) = WW^T\sigma^2_{\alpha 1}/m$ and $\text{var}(\tau_1) = I\sigma^2_{\tau 1}$, where *W* is an *n*×*m* matrix that stores standardized genotypes of *m* single nucleotide polymorphisms for n individuals, noting that WW^T is the genomic relationship matrix, and *I* is an *n*×*n* identity matrix.

Lifestyle Covariates Versus Nongenetic Determinants of Cardiovascular Traits

Classically, lifestyle covariates have been considered nongenetic determinants of cardiovascular health. In this study, lifestyle covariates and nongenetic determinants are separate concepts and serve as distinct components in our linear mixed reaction norm model (RNM). Specifically, nongenetic effects, in contrast to genetic effects, are referred to as effects of unknown factors that are not explicitly specified in our model and hence are partitioned as residual effects. On the other hand, lifestyle covariates are known factors and hence are explicitly specified in our model as modulators of genetic and nongenetic effects. The distinction is shown in Figure 1.

Heritability Estimation

We also considered the consequence of neglecting G-C and R-C interactions on heritability estimates. Specifically, we estimated heritability of each trait using 2 models, one that includes no interaction term at all, that is, the null model (that uses genomic restricted maximum likelihood [GREML] for parameter estimation), and the other, referred to as the "interaction model," that includes significant interaction terms that emerged from our primary analyses, and compared the estimates of the 2 models. To reduce computational burden, we used univariate RNMs as opposed to MRNMs. Details of univariate RNMs can be found in Data S1.

RESULTS

G-C and R-C Interactions

We had a total of 23 CVD traits, and for each trait we screened 22 available lifestyle covariates for G-C and R-C interactions. Of the 506 pairs of cardiovascular traits and lifestyle covariates, 214 yielded significant results at the 0.05 level, where the full model had a better fit than the null; after Bonferroni correction, 68 pairs remained significant (Figure 2). Of these, 34 survived a sensitivity analysis, where we applied a rank-based inverse normal transformation to all traits and refit our models. In a further investigation, we noted that a large majority of the signals that were lost after the rank-based inverse normal transformation were from traits that have large skewness and kurtosis (Figure S6). Given that rank-based inverse normal transformation can control type I error rate when the normality assumption of MRNM is violated, as shown by the simulation results (Data S2, Figures S1–S4, Tables S1 and S2), the lost signals are likely to be spurious. Hence, in the following we focus on signals that remained after the rank-based inverse normal transformation.

Of the 34 significant pairs remaining after the rank-based inverse normal transformation, 17 were covered by the UKBB, allowing replication of the analyses conducted in the ARIC Study. The majority of these signals, 14 of 17, were present in both data sets (Figure 2, right). The 3 signals lost in the replication were the modulating effects of physical activity on white blood cell count and of polyunsaturated fatty acid intake on

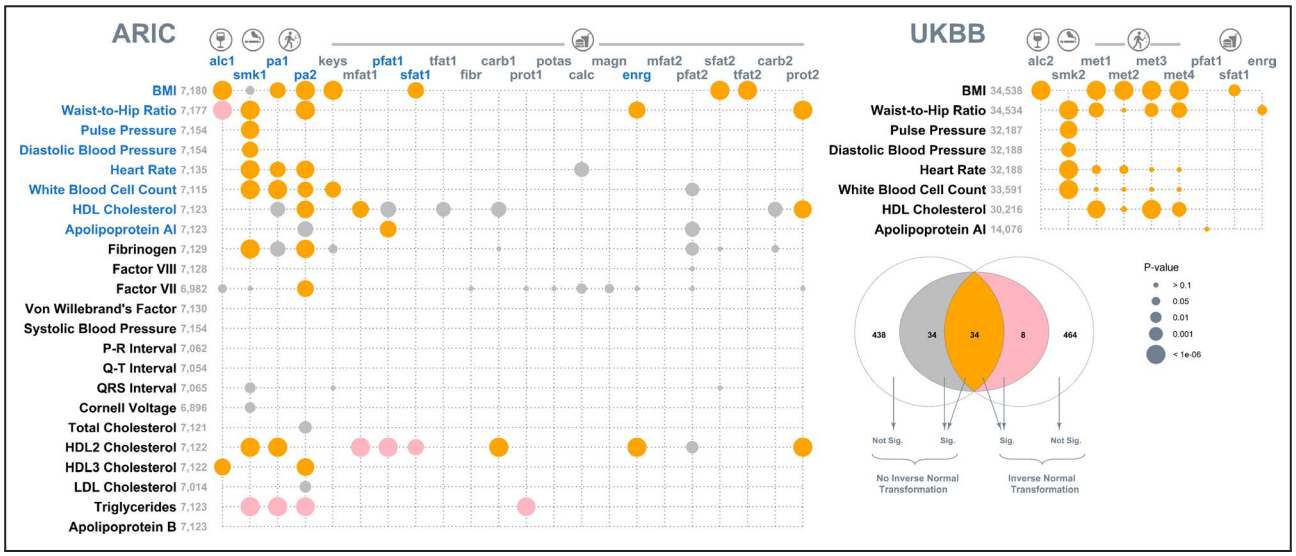


Figure 2. Bubble plot of *P* values that identify lifestyle modulation of genetic and nongenetic effects on cardiovascular traits. For each of the 23 cardiovascular traits (along the y axis) from the ARIC Study, 22 lifestyle covariates (along the x axis) were screened separately for genotype–covariate and residual–covariate interactions by comparing a multivariate reaction norm model that allows genotype–covariate and residual–covariate interactions (ie, a full model) with a null model that assumes no genotype–covariate and residual–covariate interactions (**left**). The 506 null model versus full model comparisons were repeated after a rank-based inverse normal transformation was applied to all traits for a sensitivity analysis. Signals (after Bonferroni correction) for data before and after the transformation are color coded, as detailed in the Venn diagram (**bottom right**). A total of 34 signals (in orange) remained after the sensitivity analysis. Of these remaining signals, 17 were subject to validation using the UKBB, and their corresponding traits and lifestyle covariates are highlighted in blue. The results of the UK biobank validation are shown (**top right**). For both data sets, bubbles are proportional to *P* values based on data after the rank-based inverse normal transformation. Note the exceptions to the sample size displayed for BMI versus sfat1 (N=16 257) and for waist-to-hip ratio versus enrg (N=16 254) in the UKBB because of the limited availability of dietary intake data among the selected participants. alc1 indicates alcohol intake (g/week); alc2, alcohol intake (glass and pint/week); ARIC, Atherosclerosis Risk in Communities; BMI, body mass index; calc, calcium intake (mg/d); carb1, carbohydrate intake (g/d); carb2, energy from carb1 (%kcal/d); enrg, total energy intake (kcal/d); fibr, dietary fiber intake (g/d); HDL, high-density lipoprotein; keys, keys score; LDL, low-density lipoprotein; magn, magnesium intake (mg/d); met1, summed metabolic equivalent minutes/week for all activity; met2, metabolic equivalent minutes/week for walking; met3, metabolic equivalent minutes/week for moderate activity; met4, metabolic equivalent minutes/week for vigorous activity; mfat1, monounsaturated fatty acid intake (g/d); mfat2, energy from mfat1 (%kcal/d); pa1, physical activity: leisure domain; pa2, physical activity: sports domain; pfat1, polyunsaturated fatty acid intake (g/d); pfat2, energy from pfat1 (%kcal/d); potas, potassium intake (mg/d); prot1, protein intake (g/d); prot2, energy from prot1 (%kcal/d); sfat1, saturated fatty acid intake (g/d); sfat2, energy from sfat1 (%kcal/d); Sig., significant; smk1, cigarette years of smoking; smk2, pack years adult smoking as proportion of life span exposed to smoking; tfat1, total fat intake (g/d); and tfat2, energy from tfat1 (%kcal/d); and UKBB, UK Biobank.

apolipoprotein a1. In addition, among the replicated signals, the results for physical activity varied slightly when metabolic equivalents were broken down into walking and moderate and vigorous activities, indicating that the modulating effects of physical activity may be conditional on the type of activity. The variance estimates from the full model for all signals from the ARIC Study and UKBB data sets are listed in Tables S5 and S6, respectively. In summary, our results indicate that lifestyle factors that include alcohol intake, smoking, physical activity, and dietary composition are highly relevant to interindividual variability in cardiovascular health and hence CVD risk.

For the 34 signals emerged from the ARIC Study, magnitudes of G-C and R-C interactions were further examined by expressing estimated variances of the 2 types of interactions relative to total phenotypic variances, as shown in Figure 3. Four major observations emerged. First, G-C and R-C interactions are sizeable, which can account for up to 20% of phenotypic variance, highlighting the importance of lifestyle modulation to interindividual variability in cardiovascular health. For any given trait, the larger the variance estimate of interaction effects (denoted as $\widehat{\sigma^2_{\alpha 1}}$ and $\widehat{\sigma^2_{\tau 1}}$ for G-C and R-C interaction effects, respectively), the greater the genetic or residual heterogeneity across different values of the lifestyle covariate, meaning the greater individual differences in the phenotype-lifestyle relationship; hence stronger interaction effects. Thus, variance estimates of the interaction effects can serve as measures of interaction effect size and hence are indicative of the relative importance of different lifestyle covariates to a given trait. For example, the variance estimate of G-C interactions of HDL3 cholesterol level is larger for physical activity (abbreviated as pa2) than for alcohol consumption (alc1; see Figure 3, left). Hence, phenotypic changes in HDL3 cholesterol are larger with respective to physical activity than those with respect to alcohol consumption. Similarly, the magnitude of R-C interactions is larger for fibrinogen–physical activity analysis than for fibrinogen-smoking analysis (smk1; see Figure 3, right). Standardizing

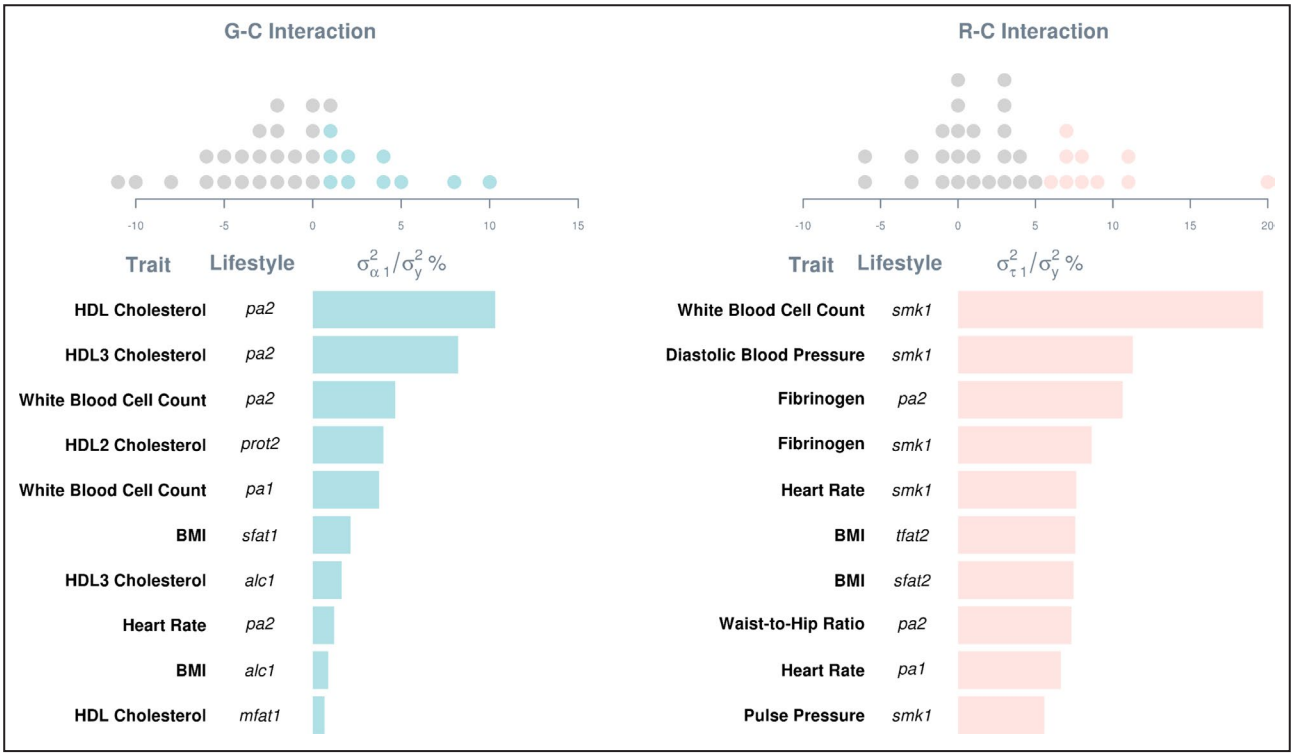


Figure 3. Variance estimates of G-C and R-C interactions as percentages of total phenotypic variance.
Estimates were derived by fitting a multivariate reaction norm model that included both G-C and R-C interactions (ie, a full model) to data without a rank-based inverse normal transformation. Dot plots on the top show distributions of estimates relative to the phenotypic variance of respective traits. Estimates are included only for signals that remained after a sensitivity analysis, where the full model was better than the null after Bonferroni correction on data after a rank-based inverse normal transformation. Top ten estimates are shown in bar charts below. *alc1* indicates alcohol intake (g/week); BMI, body mass index; G-C, genotype–covariate; HDL, high-density lipoprotein; mfat1, monounsaturated fatty acid intake (g/d); pa1, physical activity: leisure domain; pa2, physical activity: sports domain; prot2, energy from protein intake (%kcal/d); R-C, residual–covariate; $\sigma^2_{\alpha 1}$, variance of G-C interaction effects; σ^2_y, phenotypic variance; σ^2_τ, variance of R-C interaction effects; sfat1, saturated fatty acid intake (g/d); sfat2, energy from saturated fatty acid intake (%kcal/d); smk1, cigarette years of smoking; and tfat2, energy from total fat intake (%kcal/d).

cardiovascular traits also makes across-traits comparisons meaningful for the same lifestyle covariate. For example, phenotypic changes in white blood cell count are more prone to the modulation of smoking than to the modulation of fibrinogen because the variance estimate of R-C interactions is larger from the white blood cell count–smoking analysis than from the fibrinogen–smoking analysis.

Second, variance estimates of R-C interactions are in general larger than G-C interactions, indicating that lifestyle covariates play a greater role in modulating nongenetic effects on cardiovascular health than genetic effects. Third, some variance estimates can be zero or even below zero. This is not totally unexpected, though, and is within the observed range of sampling errors from analyses of the simulated data (see Data S2, Figures S1–S4, Tables S1 and S2). Lastly, we noted a strong inverse correlation between variance estimates of R-C and G-C interactions (Pearson r=−0.81). Such collinearity is likely attributed to the same covariate being used for estimating G-C and R-C interactions. Similar observations were noted in each replicate of simulated data, yet both variance estimates of G-C and R-C interactions were unbiased (see Data S2, Figures S1–S4, Tables S1 and S2). Thus, despite the collinearity between variance estimates, estimation accuracy did not appear to be adversely affected. It is noted that the statistical power to separate G-C and R-C interactions can be low, and parameter estimates from models including only G-C or R-C interaction (referred to as "G-C only" and "R-C only" models) can be biased as shown in simulations (see Data S2, Figures S1–S4, Tables S1 and S2). Consequently, only the null model versus the full model comparison was chosen to indicate lifestyle modulation. Nonetheless, we compared nested models, that is, a G-C only model and a R-C only model, with the full model to assess R-C interaction that is orthogonal to G-C interaction and G-C interaction that is orthogonal to R-C interaction, respectively (Table S7 for the ARIC Study and Table S8 for UKBB).

For the 14 signals that were first discovered in the ARIC Study and replicated in UKBB, we compared variance estimates of G-C and R-C interaction effects across the 2 data sets (Tables S5 and S6) and noted some similarities. Physical activity modulates both genetic and nongenetic effects on heart rate and BMI. It also modulates genetic effects on HDL cholesterol level and nongenetic effects on waist-to-hip ratio. Alcohol consumption modulates both genetic and nongenetic effects on BMI, whereas smoking modulates nongenetic effects on heart rate, pulse pressure, and white blood cell count. In addition, saturated fat intake modulates genetic effects on BMI, and total daily energy intake modulates nongenetic effects on waist-to-hip ratio.

The presence of G-C and R-C interactions indicates heterogeneity of genetic and residual variance–covariance structures with respect to lifestyle covariates,[39] which are depicted in Figure S7 for G-C interactions and in Figure S8 for R-C interactions. To explicitly illustrate G-C interactions, for each of the 8 traits with the largest variance estimates of G-C interaction, we stratified observations into 3 groups—top, middle, and bottom—according to the per-individual estimate of G-C interaction effects, denoted as $\hat{\alpha}_1$ (via the best linear unbiased prediction[40,41]). It is important to note that α_1 in our model indicates the direction and effect size of the G-C interaction effect for each individual, and it is assumed to follow a normal distribution with the mean zero. For each trait, we defined the 3 groups as having an $\hat{\alpha}_1$ below the 20th percentile of the sample (bottom), between the 40th and 60th percentiles (middle), and above the 80th percentile (top), respectively, and plotted their phenotypic estimates, that is, $\hat{\alpha}_0 + \cdot\hat{\alpha}_1$, given their standardized values on the relevant lifestyle covariate c (Figure 4). Irrespective of the trait, 3 groups showed distinct trajectories of phenotypic changes with respect to lifestyle covariate. Thus, significant G-C interactions indicated that there exist genetically distinct subpopulations with different phenotype–lifestyle relationships, and hence per-individual estimates of G-C interactions inform individual differences, by genetic predisposition, in the extent to which one may benefit from lifestyle changes.

Importantly, G-C interactions potentially have important clinical relevance. To illustrate, we use the HDL cholesterol–physical activity analysis (with the largest variance estimates of G-C interaction effects) as an example. Figure 5A shows the predicted trajectories of phenotypic changes in HDL cholesterol level as a function of physical activity for individuals stratified by the percentile group of estimated G-C interaction (80–85%, 85–90%, 90–95%, and 95–100%). For people with a G-C interaction estimate that falls between 95% and 100% of the sample, every standard unit increase in physical activity is associated with an average increase in HDL cholesterol level by 0.3 standard unit. This is about 4 times greater than that (0.07 standard unit) for individuals with an interaction estimate that falls between 80% to 85% of the sample. Given the association between HDL cholesterol increase and CVD risk reduction (eg, ref. 42), an increase in physical activity would be most beneficial to individuals in the group that falls between 95% to 100%.

Importantly, the per-individual estimate of α_1 that we used for group stratification is an aggregate of G-C interactions over common SNPs of the whole genome, hence a genome-wide estimate of G-C interaction. Therefore, individual differences in $\hat{\alpha}_1$ would reflect systematic genetic variation, which would

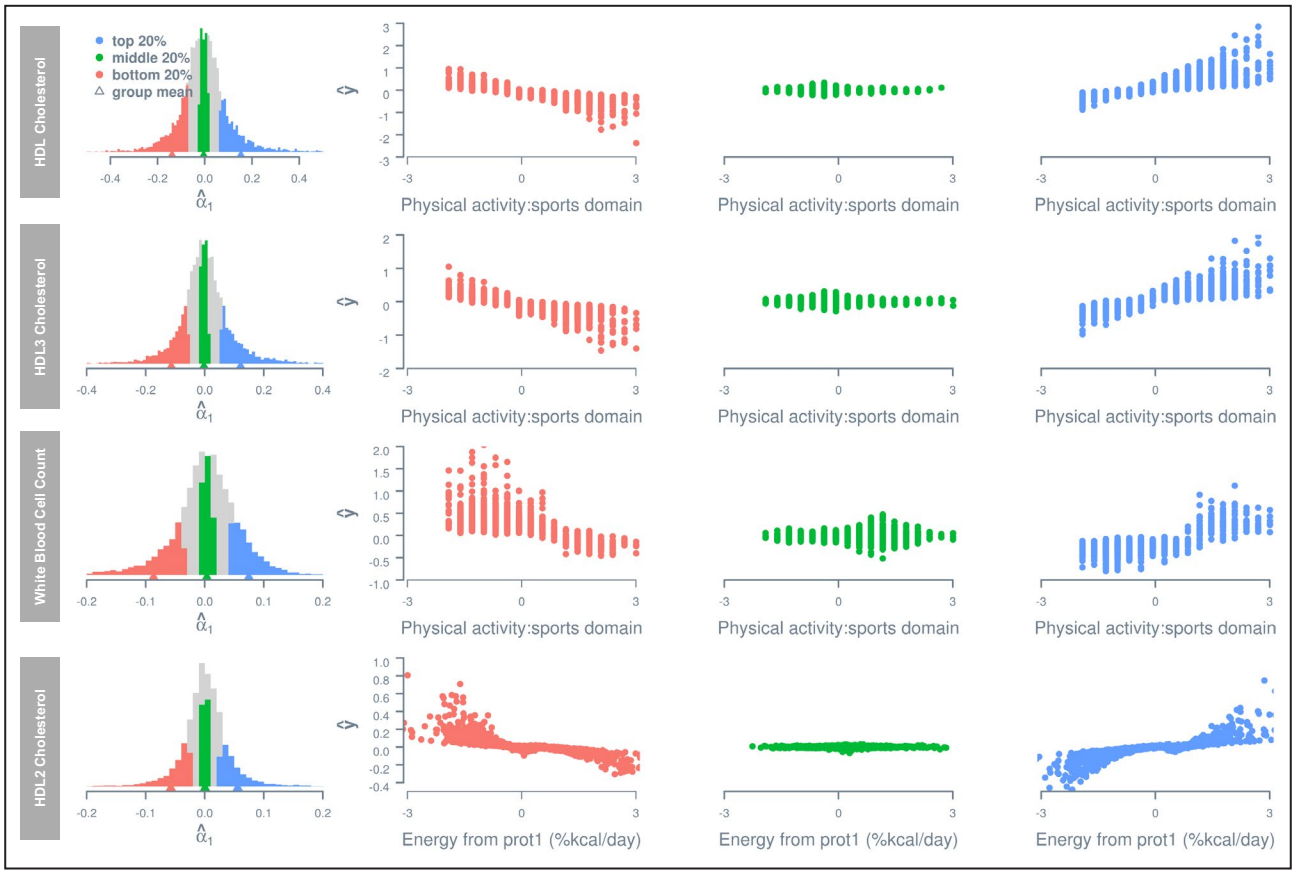


Figure 4. Estimated phenotypes as a function of lifestyle covariates for groups stratified by per-individual estimate of genotype–covariate interaction effect.
Histograms on the left show distributions of per-individual estimates of a genotype–covariate interaction effect, that is, α_1. The estimated phenotype of any given individual *i* is computed using the equation $\hat{y}_i = \hat{\alpha}_{0i} + c_i\,\hat{\alpha}_{1i}$, where *c* denotes the lifestyle covariate value recorded for *i*, and $\hat{\alpha}_{0i}$ and $\hat{\alpha}_{1i}$ denote the estimated main genetic effect and genotype–covariate interaction effect for *i*. Only the first 4 traits with the largest variance estimate of genotype–covariate interaction effects are shown. All phenotypes and lifestyle covariates are standardized. HDL, high-density lipoprotein; and prot1, protein intake (g/d).

be the most pronounced between the 2 extreme groups, that is, top and bottom. A further exploration on pairwise genomic relationships for individuals within and between the 2 extreme groups revealed that the average genomic relationship within each group is greater than the grand average relationship of the entire data set, but the average between-group relationship is less than the grand average (Table S9). That is, compared with 2 randomly chosen individuals, a pair of within-group individuals is on average more genetically similar, but a pair of between-group individuals is on average more genetically distant. This observation holds for all 8 analyses with the largest variance estimates of G-C interaction (Table S9). Thus, the 2 extreme groups for these analyses in fact have systematic genetic differences.

To explicitly illustrate R-C interactions, for each of the 8 traits with the largest variance estimates of R-C interaction, we stratified participants into top, middle, and bottom groups according to per-individual estimate of R-C interaction effects, denoted as $\hat{\tau}_1$, in the same way as for G-C interaction. Figure 6 shows estimated phenotypes, that is, $\hat{\alpha}_0 + c \cdot \hat{\tau}_1$, given standardized values on the relevant lifestyle covariate c for the 3 groups. Similar to G-C interaction, the 3 groups show different phenotypic changes with increasing lifestyle covariate values. Thus, similar to G-C interactions, R-C interactions indicate the presence of distinct subpopulations with different phenotype–lifestyle relationships, and hence per-individual estimates of R-C interactions inform individual differences in the extent to which one may benefit from lifestyle changes, which are relevant to clinicians and health professionals equally as G-C interaction estimates.

Figure 5B shows the predicted trajectories of phenotypic changes in white blood cell count as a function of cigarette smoking for individuals stratified by the percentile group of estimated R-C interaction (80–85%, 85–90%, 90–95%, and 95–100%). For people with a R-C

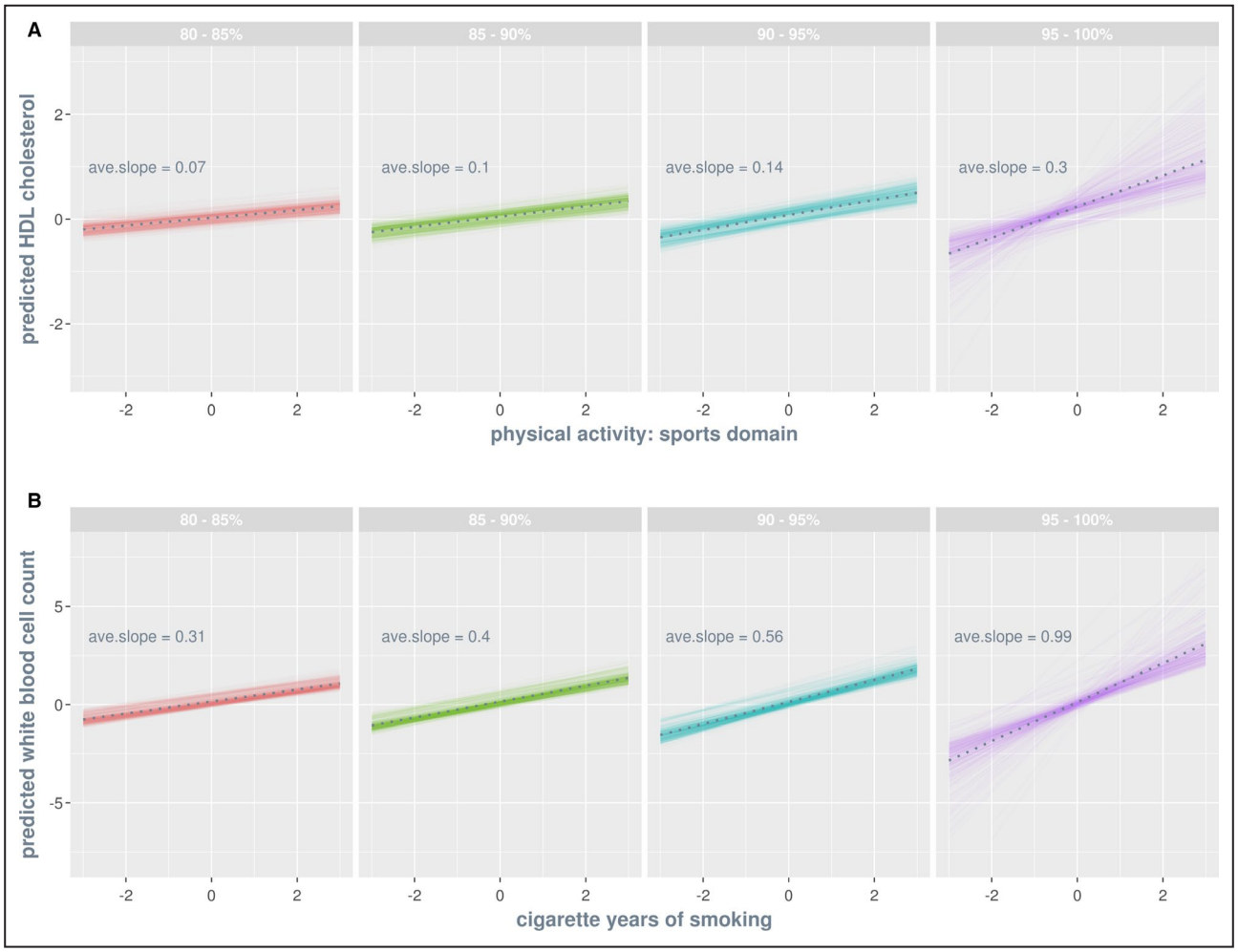


Figure 5. Predicted trajectories of phenotypic changes as a function of lifestyle covariate by percentile group of estimated genotype–covariate interaction effects (A) and of estimated residual-covariate interaction effects (B).
Percentile groups are color coded, and within each group faint lines are individuals and the dotted line is the group average. Both traits (shown on the *y* axis) and lifestyle covariates (shown on the *x* axis) are standardized such that the slope of a given trajectory indicates the number of standard unit change in the phenotype of a trait per standard unit change in a lifestyle covariate. **A**, The HDL cholesterol–physical activity analysis is chosen for illustration because it has the largest variance estimate of genotype–covariate interactions. The predicted phenotypes for an individual *i* are computed by substituting lifestyle covariate values, between –3 and 3, into the equation $\hat{\alpha}_{0i} + c_i\hat{\alpha}_{1i}$, where *c* denotes lifestyle covariate, $\hat{\alpha}_{0i}$ and $\hat{\alpha}_{0i}$ denote the estimated main genetic effect and genotype–covariate interaction effect for *i*, respectively. **B**, The white blood cell count–smoking analysis is chosen for illustration because it is the largest variance estimate of residual-covariate interactions. The predicted phenotypes for individual *i* are computed by substituting lifestyle covariate values, between –3 and 3, into the equation $\hat{\alpha}_{0i} + c_i\hat{\tau}_{1i}$, where $\hat{\tau}_i$ denotes the estimated residual-covariate interaction effect for *i*. ave., average; and HDL, high-density lipoprotein.

interaction estimate that falls between 95% and 100% of the sample, every standard unit decrease in smoking is associated with an average reduction in white blood cell count by 0.99 standard unit. This is about 3 times greater than that (0.31 standard unit) for individuals with an interaction estimate that falls between 80% to 85% of the sample. Given the association between white blood cell count decrease and CVD risk reduction (eg, ref. 43), whereas all percentile groups would benefit from a reduction in smoking, the most benefit would be evident for individuals in the 95% to 100% group.

Although a genuine R-C interaction can be unbiasedly estimated as shown in the simulation (see Data S2, Figures S1–S4, Tables S1 and S2), the models used in this study do not inform how individual differences in $\hat{\tau}_1$ arise because the fitted variance–covariance structure for residual effects is an identity matrix (see Data S1). However, this problem will no longer exist in a repeated-measures design or if a nonidentity matrix is fitted for the variance–covariance structure of residual effects.[10]

Heritability

We showed previously that lifestyle modulation of genetic and nongenetic effects, in forms of G-C and R-C

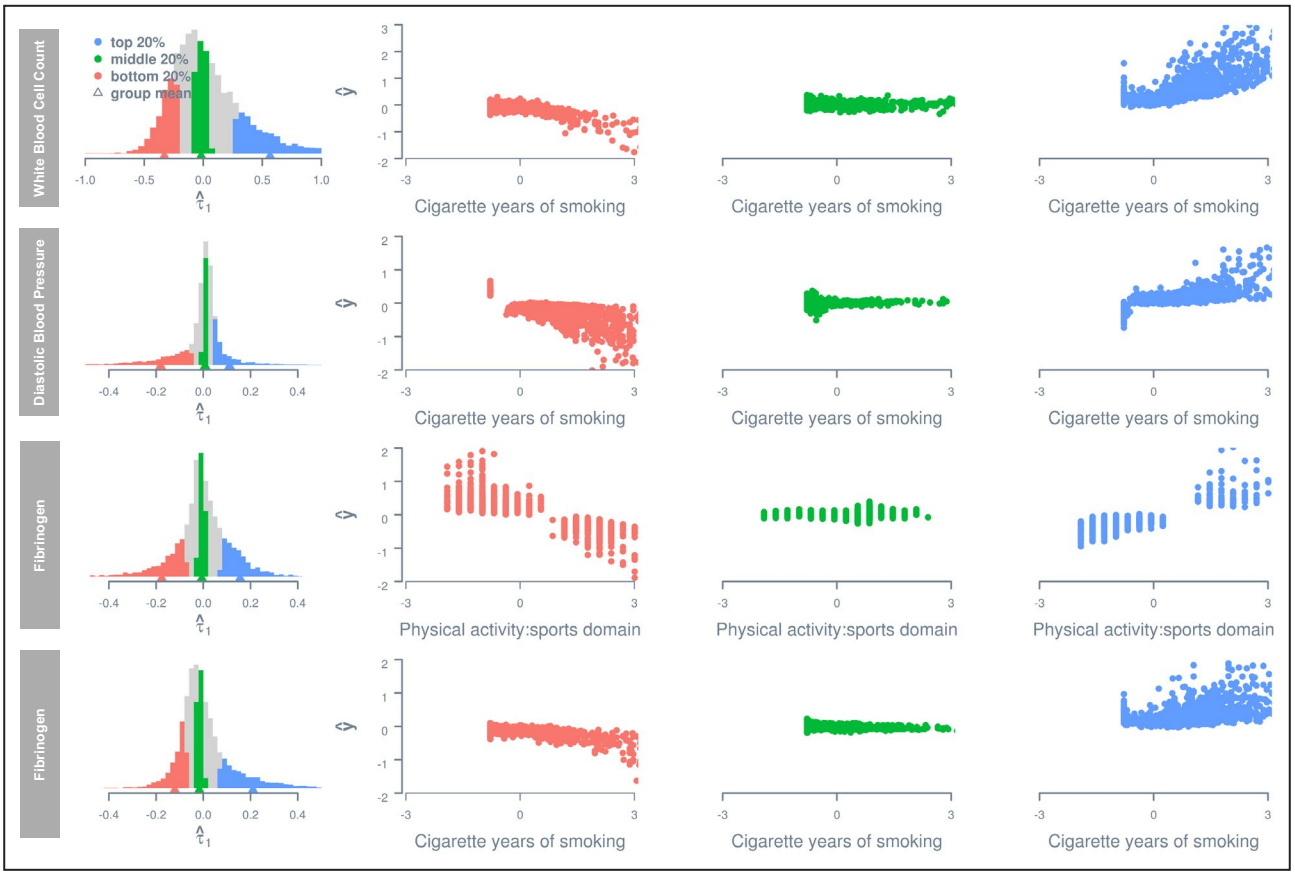


Figure 6. Estimated phenotypes as a function of lifestyle covariates for groups stratified by per-individual estimates of residual–covariate interaction.

Histograms on the left show distributions of per-individual estimates of a residual–covariate interaction effect, τ_1. The estimated phenotype of any given individual *i* is computed using the equation $\hat{y}_i = \hat{\alpha}_{0i} + c_i\hat{\tau}_{1i}$, where *c* denotes the lifestyle covariate value recorded for *i*, $\hat{\alpha}_{0i}$ and $\hat{\tau}_{1i}$ denote the estimated main genetic effect and residual–covariate interaction effect for *i*, respectively. Only the first 4 traits with the largest variance estimate of residual–covariate interaction effects are shown. All phenotypes and lifestyle covariates are standardized.

interactions, is ubiquitous and sizable for cardiovascular traits. To highlight the importance of incorporating lifestyle modulation when estimating trait heritability, we compared SNP heritability estimates from 2 univariate RNMs (see Methods for details), one without any interaction terms (ie, a null model) and the other with interaction terms (ie, an interaction model) based on results from the MRNMs shown previously (Figure 2). Null model estimates are essentially equivalent to conventional univariate GREML estimates; hence they are referred to as GREML estimates thereafter. As a contrast, interaction model estimates are thereafter referred to as RNM estimates. Figure 7 is a scatter plot of estimates from both the ARIC Study and UKBB data sets. If GREML and RNM estimates are identical, they are expected to align perfectly along the diagonal line. We found that estimates from the interaction model were, on average, slightly yet systematically larger than estimates from the null model (single-sided paired t=2.35, df=17, P=0.015). Thus, our results support the idea that phenotypic plasticity[39] can explain some missing heritability (eg, ref. 44).

Given that heritability is a function of genetic and residual variance, we further investigated the reason behind larger RNM heritability estimates by comparing GREML and RNM estimates of genetic and residual variance from both the ARIC Study and UKBB data sets (Figure 8). On average, GREML and RNM estimates of genetic variance were not significantly different, but GREML estimates of residual variance were significantly larger than RNM estimates (see mean and 95% CI in Figure 8; 2-sided 1-sample t=4.15, df=17, P=6.7×10^{-4}). However, the results from the ARIC Study and UKBB data sets are somewhat different. In particular, some GREML estimates of genetic variance tend to be underestimated for the ARIC Study data, which is not evident for the UKBB data. This is likely attributed to the smaller sample size of the ARIC Study data than the UKBB data, which inevitably results in larger sampling errors for the estimates from the ARIC Study. Nonetheless, our results indicate that G-C and R-C interactions are primarily hidden in residual variance estimates in the null model; when they

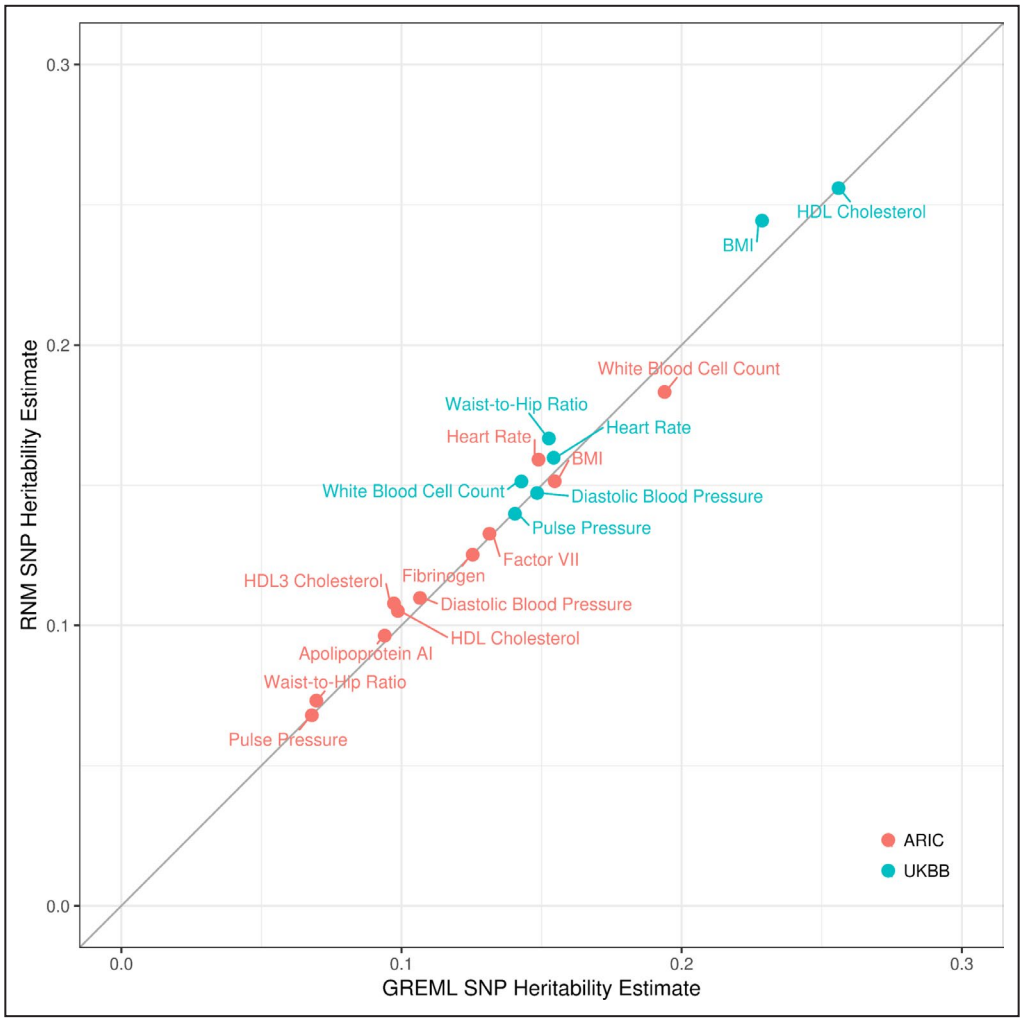


Figure 7. SNP heritability estimates based on GREML and RNM.
GREML and RNM estimates were derived by fitting a univariate reaction norm model including no interaction term (ie, null model) and one including 1 or more interaction terms (ie, interaction model), respectively. The diagonal is included to highlight the impact of neglecting interaction terms on SNP heritability estimates. Deviations above the diagonal indicate larger RNM estimates relative to GREML estimates. Note HDL2 cholesterol is excluded because of negative variance estimates. ARIC, Atherosclerosis Risk in Communities; BMI, body mass index; GREML, genomic restricted maximum likelihood; HDL, high-density lipoprotein; RNM, reaction norm models; SNP, single nucleotide polymorphism; and UKBB, UK Biobank.

are explicitly estimated in an interaction model, residual variance estimates can be substantially reduced, thereby yielding higher SNP heritability compared with when these components are neglected. This is in line with our previous observation that residual variance is overestimated when fitting a null model to simulated data with genuine G-C or R-C interaction.[10]

DISCUSSION

In this study, we used a novel linear mixed model to detect and estimate components of genetic and nongenetic variance that change with respect to modifiable lifestyle covariates, termed as G-C and R-C interactions, in the context of cardiovascular health. Using simulations, we showed that for a sample size of ≈7500 observations, our method has sufficient statistical power to detect genuine G-C and R-C interactions while keeping the false positive rate controlled. Applying our method to real data, for each of 23 cardiovascular traits selected from the ARIC Study data set, we screened for G-C and R-C interactions using 22 available lifestyle covariates that covered smoking, alcohol intake, physical activity, and dietary composition. G-C and R-C interactions were found to be ubiquitous among cardiovascular health related traits, and for some traits, estimates

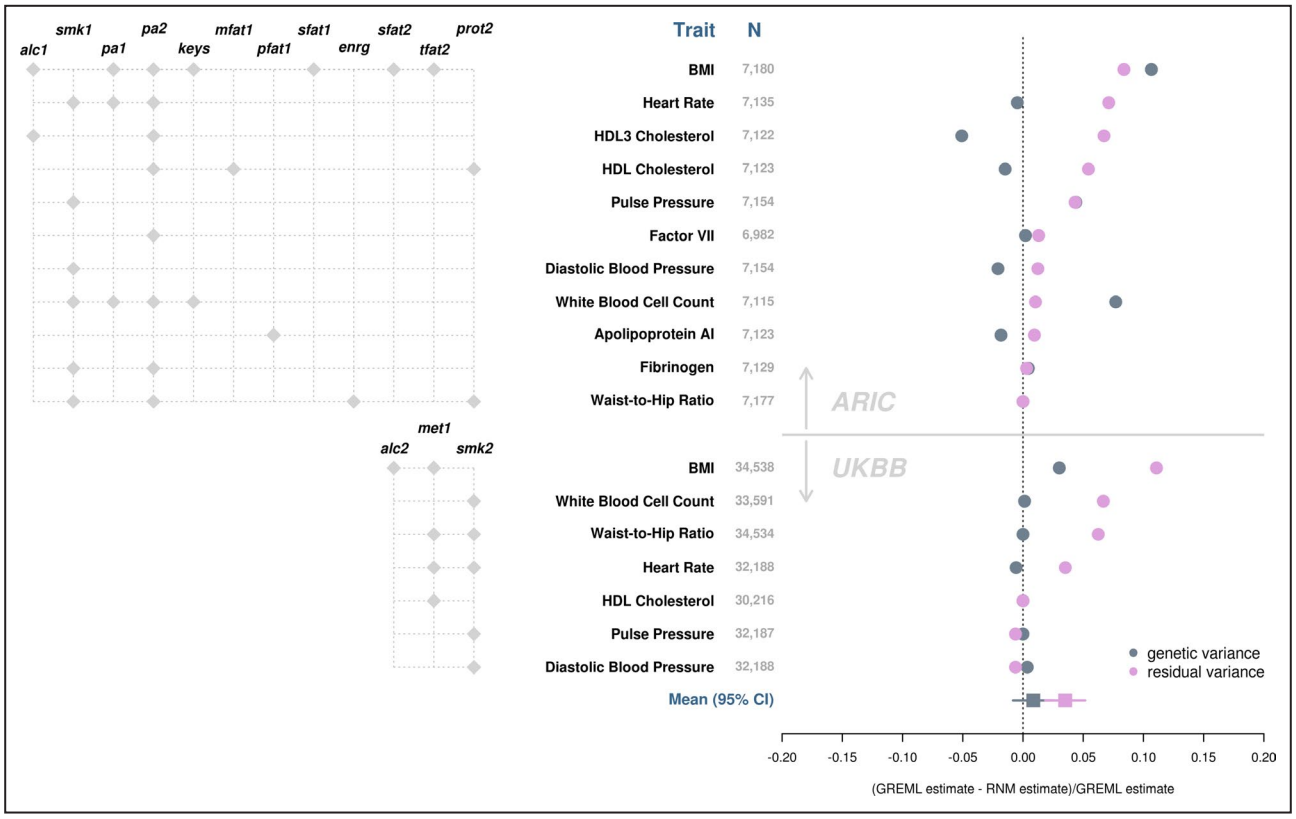


Figure 8. Genetic and residual variance estimates based on GREML and RNM.
GREML and RNM estimates were derived by fitting a univariate RNM that included no interaction term (ie, null model) and one that included interaction terms (ie, interaction model), respectively. Lifestyle covariate(s) included in the interaction model are specified (**left**). Changes in genetic and residual variance estimates from the interaction model (ie, RNM estimates) relative to their respective estimates from the null model (ie, GREML estimates) are shown to highlight the impact of neglecting interaction terms (**right**). Deviations below 0 indicate underestimation by GREML, whereas deviations above 0 indicate overestimation by GREML. Traits are presented in the decreasing order of deviations for residual variance. Note changes in genetic variance estimates for pulse pressure and waist-to-hip ratio in the ARIC Study and for HDL cholesterol in the UKBB are obscured in the plot by data points for residual variance. HDL2 cholesterol in ARIC was excluded because of the negative heritability estimates. alc1 indicates alcohol intake (g/week); alc2, alcohol intake (glass and pint/week); ARIC, Atherosclerosis Risk in Communities; BMI, body mass index; enrg, total energy intake (kcal/d); keys, keys score; GREML, genomic restricted maximum likelihood; HDL, high-density lipoprotein; met1, summed MET minutes/week for all activity; mfat1, monounsaturated fatty acid intake (g/d); pa1, Physical activity: leisure domain; pa2, physical activity: sports domain; pfat1, polyunsaturated fatty acid intake (kcal/d); prot2, energy from protein intake (%kcal/d); RNM, reaction norm model; sfat1, saturated fatty acid intake (g/d); sfat2, energy from saturated fatty acid intake (%kcal/d); smk1, cigarette years of smoking; smk2, pack years adult smoking as proportion of life span exposed to smoking; tfat2, energy from total fat intake (%kcal/d); and UKBB, UK Biobank.

were relatively large, accounting for up to 20% of the total phenotypic variance.

Among the 14 signals replicated in the UKBB, physical activity was found to alter both genetic and nongenetic effects on heart rate and BMI; genetic effects on HDL cholesterol level and nongenetic effects on waist-to-hip ratio. Alcohol consumption altered both genetic and nongenetic effects on BMI, whereas smoking altered nongenetic effects on heart rate, pulse pressure, and white blood cell count. In addition, saturated fat intake modified genetic effects on BMI, and total daily energy intake modified nongenetic effects on waist-to-hip ratio. To explicitly illustrate G-C and R-C interactions, we stratified individuals according to the per-individual estimate of G-C and R-C interactions and showed that genetic and residual effects could take on different directions across groups. Although we did not identify any literature in the context of cardiovascular traits that examined R-C interaction, the evidence of G-C interaction in our studies is consistent with the previous literature (eg, refs. 7,9,13,16–20). Our study is novel in that G-C interactions were estimated using common SNPs across the entire genome, which are in contrast to estimates based on a single or a limited number of SNPs with large phenotypic effects in past studies.

Given the prevalence of lifestyle modulating effects, we also examined any potential consequence of neglecting these effects on SNP heritability estimates.

Such negligence reduced SNP heritability estimates by a small yet significant amount. This reduction is primarily attributed to the overestimation of residual variance. Yet genetic variance estimates are relatively robust to the negligence of significant lifestyle modulation. Our results suggest that current SNP heritability estimates for cardiovascular health-related outcomes, which commonly do not take into account modulating effects of lifestyle covariates, are likely underestimated.

Currently, several other approaches to G-C interaction exist in the literature, and our approach is unique in several important ways. Compared with a fixed effects model approach (eg, ref. 45), a mixed-model approach such as ours could account for genetic covariance among individuals. Compared with StructLMM,[8] which is a linear mixed-model approach that examines G-C interaction for 1 SNP at a time, our approach estimates the G-C interaction aggregated over SNPs for the entire genome, thereby providing genome-wide estimates of G-C interaction. The whole-genome approach to G-C interaction also sets this study apart from conventional G-C interaction studies using a candidate gene approach that focus on only a few genetic variants with large phenotypic effects (eg, refs. 13–15). In addition, our approach extends other whole-genome approaches[7,33,46] by allowing continuous, as opposed to categorical, lifestyle covariates to be modeled and by simultaneously modeling G-C and R-C interactions.

The prevalence of the sizable G-C and R-C interaction effects shown in our study not only reinforces the relevance of existing lifestyle-focused prevention programs for CVD prevention but also suggests that promoting lifestyle changes in a single direction may be ineffective or even inappropriate for some subpopulations. Instead, to most effectively reduce genetic and nongenetic predispositions to unfavorable cardiovascular phenotypes, lifestyle-focused interventions should be tailored to the individual on the basis of his or her relevant genetic and nongenetic information, supporting the rise of precision medicine in CVD to individualize treatments and preventions rather than assuming all individuals share a common pathophenotype.[47,48]

Of note, the variance–covariate structure fitted for the genetic effect in our MRNMs is a nonidentity matrix constructed using genetic information, that is, a genomic relationship matrix. In effect, the SNP best-linear unbiased predictions derived from MRNMs can be used to predict how a person's genetic risk would change with respect to a chosen lifestyle covariate given his or her genetic information. In contrast, the variance–covariance structure fitted for the residual effect in our MRNMs is an identity matrix. Consequently, R-C interactions estimated by our models have little use in the prediction of phenotypes. Further development of MRNMs that incorporate a relationship matrix based on factors underlying residual variations, that is, a nonidentity matrix analogous to a genomic relationship matrix, would be useful for the prediction, and it is currently under way in a separate study.

As for other approaches to G-C interaction for observational studies, the modulating effects of lifestyle covariates found in this study do not imply causality. Although randomized controlled trials are the gold standard, further studies using genetic methods such as Mendelian randomization can help determine causal influences. Furthermore, the MRNMs used in this article are a specific case of the more general MRNMs (see ref. 10), where genetic and residual effects are expanded to the first order of the chosen lifestyle covariate. Higher order expansions may be necessary and could be employed in future studies where the variance–covariance structure for residual effects is a nonidentity matrix. However, increasing model complexity also increases notably the sample size requirement for robust estimation of model parameters. It should also be noted that the sample size in the ARIC Study for our primary analyses is relatively small (6896–7180 participants), leading to less precise parameter estimation compared with the UKBB validation analyses. This may explain some discrepancies observed in the model estimates between the 2 data sets. Smoking, for example, was shown to modulate cardiovascular health in both data sets, but the modulations manifested primarily as R-C interactions in the ARIC Study analyses but as G-C interactions in UKBB analyses (see Tables S5 and S6). Independent data sets are required to determine the nature of the modulation effects of smoking on cardiovascular traits. Finally, in this article we only considered intermediate cardiovascular traits that are continuous in nature. The development of valid MRNMs for binary outcome is currently under way. Future applications of these MRNMs would help identify modulating lifestyle covariates that are directly relevant to CVD outcomes.

In summary, we found strong modulations from lifestyle covariates, including smoking, alcohol intake, physical activity, and dietary composition, for genetic and residual effects on phenotypes that are known to associate with CVDs. To illustrate these interactions, we showed that genetic and residuals effects—which may be interpreted as genetic and nongenetic predisposition to CVD health risk, respectively—could change with respect to lifestyle change in different directions for different individuals. Our findings, therefore, reinforce the relevance of lifestyle changes to cardiovascular health and highlight the need for individual considerations when designing lifestyle intervention programs to effectively reduce genetic and nongenetic predispositions to unfavorable cardiovascular phenotypes. Future investigations into specific genetic and nongenetic factors that give rise to individual

differences in CVD health risk trajectories with respect to lifestyle changes are well warranted.

ARTICLE INFORMATION

Received January 2, 2020; accepted March 13, 2020.

Affiliations

From the Australian Centre for Precision Health, University of South Australia, Adelaide, South Australia, Australia (X.Z., K.C.-C., E.H., S.H.L.); School of Environmental and Rural Science, University of New England, Armidale, New South Wales, Australia (J.v.d.W., G.N.); Institute for Molecular Bioscience (G.N.) and Queensland Brain Institute (J.M.), University of Queensland, Brisbane, Queensland, Australia; Queensland Centre for Mental Health Research, The Park Centre for Mental Health, Wacol, Queensland, Australia (J.M.); South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia (X.Z., E.H., S.H.L.).

Acknowledgments

The authors would like to thank staff and participants of the Atherosclerosis Risk in Communities Study and the UK Biobank for their important contributions. Work was performed using computational resources provided by the Australian Government through Raijin under the National Computational Merit Allocation Scheme.

Sources of Funding

This research is supported by the Australian National Health and Medical Research Council (1080157) and the Australian Research Council (DP160102126, DP190100766, FT160100229). The Atherosclerosis Risk in Communities Study has been funded in whole or in part with federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (under contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I). The UK Biobank is funded by the UK Department of Health, the Medical Research Council, the Scottish Executive, and the Wellcome Trust Medical Research Charity.

Disclosures

None.

Supplementary Materials

Data S1–S3
Tables S1–S9
Figures S1–S8

REFERENCES

1. World Health Organization. Fact sheet: cardiovascular diseases. https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Published May 17, 2017. Accessed February 20, 2020.
2. Waken RJ, de las Fuentes L, Rao DC. A review of the genetics of hypertension with a focus on gene-environment interactions. *Curr Hypertens Rep*. 2017;19:23.
3. Namboodiri KK, Kaplan EB, Heuch I, Elston RC, Green PP, Rao DC, Laskarzewski P, Glueck CJ, Rifkind BM, Skolnick MH. The collaborative lipid research clinics family study: biological and cultural determinants of familial resemblance for plasma lipids and lipoproteins. *Genet Epidemiol*. 1985;2:227–254.
4. Freeman MS, Mansfield MW, Barrett JH, Grant PJ. Genetic contribution to circulating levels of hemostatic factors in healthy families with effects of known genetic polymorphisms on heritability. *Arterioscler Thromb Vasc Biol*. 2002;22:506–510.
5. Vossen CY, Callas PW, Hasstedt SJ, Long GL, Rosendaal FR, Bovill EG. A genetic basis for the interrelation of coagulation factors. *J Thromb Haemost*. 2007;5:1930–1935.
6. Nowak-Göttl U, Langer C, Bergs S, Thedieck S, Sträter R, Stoll M. Genetics of hemostasis: differential effects of heritability and household components influencing lipid concentrations and clotting factor levels in 282 pediatric stroke families. *Environ Health Perspect*. 2008;116:839.
7. Robinson MR, English G, Moser G, Lloyd-Jones LR, Triplett MA, Zhu Z, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H, The LifeLines Cohort Study, et al. Genotype–covariate interaction effects and the heritability of adult body mass index. *Nat Genet*. 2017;49:1174.
8. Moore R, Casale FP, Jan Bonder M, Horta D, Heijmans BT, C.'t Hoen PA, van Meurs J, Isaacs A, Jansen R, Franke L, et al. A linear mixed-model approach to study multivariate gene—environment interactions. *Nat Genet*. 2019;51:180–186.
9. Young AI, Wauthier F, Donnelly P. Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. *Nat Commun*. 2016;7:12724.
10. Ni G, van der Werf J, Zhou X, Hyppönen E, Wray NR, Lee SH. Genotype–covariate correlation and interaction disentangled by a whole-genome multivariate reaction norm model. *Nat Commun*. 2019;10:2239.
11. Maher B. Personal genomes: the case of the missing heritability. *Nature*. 2008;456:18–21.
12. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747.
13. Hindy G, Wiberg F, Almgren P, Melander O, Orho-Melander M. Polygenic risk score for coronary heart disease modifies the elevated risk by cigarette smoking for disease incidence. *Circ Genom Precis Med*. 2018;11:e001856.
14. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, Chasman DI, Baber U, Mehran R, Rader DJ, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med*. 2016;375:2349–2358.
15. Rutten-Jacobs LC, Larsson SC, Malik R, Rannikmäe K, MEGASTROKE Consortium, International Stroke Genetics Consortium, Sudlow CL, Dichgans M, Markus HS, Traylor M. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants. *BMJ*. 2018;363:k4168.
16. Corella D, Peloso G, Arnett DK, Demissie S, Cupples LA, Tucker K, Lai C-Q, Parnell LD, Coltell O, Lee Y-C. Apoa2, dietary fat, and body mass index: replication of a gene-diet interaction in 3 independent populations. *Arch Intern Med*. 2009;169:1897–1906.
17. Latella MC, Di Castelnuovo A, De Lorgeril M, Arnout J, Cappuccio FP, Krogh V, Siani A, Van Dongen M, Donati MB, De Gaetano G. Genetic variation of alcohol dehydrogenase type 1C (adh1c), alcohol consumption, and metabolic cardiovascular risk factors: results from the immidiet study. *Atherosclerosis*. 2009;207:284–290.
18. Bernstein MS, Costanza MC, James RW, Morris MA, Cambien F, Raoux S, Morabia A. Physical activity may modulate effects of APOE genotype on lipid profile. *Arterioscler Thromb Vasc Biol*. 2002;22:133–140.
19. Corella D, Guill M, Portol O, Sabater A, Cortina S, Ordovas JM. Environmental factors modulate the effect of the APOE genetic polymorphism on plasma lipid concentrations: ecogenetic studies in a mediterranean Spanish population. *Metab, Clin Exp*. 2001;50:936–944.
20. Ruaño G, Seip RL, Windemuth A, Zöllner S, Tsongalis GJ, Ordovas J, Otvos J, Bilbie C, Miles M, Zoeller R. Apolipoprotein A1 genotype affects the change in high density lipoprotein cholesterol subfractions with exercise training. *Atherosclerosis*. 2006;185:65–69.
21. The ARIC Investigators. The atherosclerosis risk in community (ARIC) study: design and objectwes. *Am J Epidemiol*. 1989;129:687–702.
22. Papp A, Hatzakis H, Bracey A, Wu K. ARIC hemostasis study—I. Development of a blood collection and processing system suitable for multicenter hemostatic studies. *Thromb Haemost*. 1989;61:015–019.
23. Seaman CD, George KM, Ragni M, Folsom AR. Association of Von Willebrand factor deficiency with prevalent cardiovascular disease and asymptomatic carotid atherosclerosis: the atherosclerosis risk in communities study. *Thromb Res*. 2016;144:236.
24. Folsom AR, Nieto FJ, Sorlie P, Chambless LE, Graham DY. Helicobacter pylori seropositivity and coronary heart disease incidence. *Circulation*. 1998;98:845–850.
25. O'Neal WT, Singleton MJ, Roberts JD, Tereshchenko LG, Sotoodehnia N, Chen LY, Marcus GM, Soliman EZ. Association between QT-interval components and sudden cardiac death: the ARIC study (Atherosclerosis Risk in Communities). *Circ Arrhythm Electrophysiol*. 2017;10:e005485.
26. Decker WW, Prina LD, Smars PA, Boggust AJ, Zinsmeister AR, Kopecky SL. Continuous 12-lead electrocardiographic monitoring in an emergency department chest pain unit: an assessment of potential clinical effect. *Ann Emerg Med*. 2003;41:342–351.

27. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol*. 1985;122:51–65.
28. Shekelle RB, Shryock AM, Paul O, Lepper M, Stamler J, Liu S, Raynor WJ Jr. Diet, serum cholesterol, and death from coronary heart disease: the western electric study. *N Engl J Med*. 1981;304:65–70.
29. Stamler J, Elliott P, Appel L, Chan Q, Buzzard M, Dennis B, Dyer AR, Elmer P, Greenland P, Jones D. Higher blood pressure in middle-aged American adults with less education—role of multiple dietary factors: the intermap study. *J Hum Hypertens*. 2003;17:655.
30. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr*. 1982;36:936–942.
31. Folsom AR, Arnett DK, Hutchinson RG, Liao F, Clegg LX, Cooper LS. Physical activity and incidence of coronary heart disease in middle-aged women and men. *Med Sci Sports Exerc*. 1997;29:901–909.
32. Bell EJ, Lutsey PL, Windham BG, Folsom AR. Physical activity and cardiovascular disease in African Americans in ARIC. *Med Sci Sports Exerc*. 2013;45:901.
33. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW. Common SNPS explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565.
34. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.
35. Cassidy S, Chau JY, Catt M, Bauman A, Trenell MI. Cross-sectional study of diet, physical activity, television viewing and sleep duration in 233 110 adults from the UK Biobank; the behavioural phenotype of cardiovascular disease and type 2 diabetes. *BMJ Open*. 2016;6:e010038.
36. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45:984.
37. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, Bergen SE, Collins AL, Crowley JJ, Fromer M. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet*. 2013;45:1150.
38. Lee SH, Yang J, Chen G-B, Ripke S, Stahl EA, Hultman CM, Sklar P, Visscher PM, Sullivan PF, Goddard ME. Estimation of SNP heritability from dense genotype data. *Am J Hum Genet*. 2013;93:1151–1155.
39. Lynch M, Walsh B. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer; 1998.
40. Clark SA, van der Werf J. Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values. *Methods Mol Biol*. 2013;1019:321–330.
41. Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics*. 1975;31:423–447.
42. Gordon DJ, Probstfield JL, Garrison RJ, Neaton JD, Castelli WP, Knoke JD, Jacobs DR Jr, Bangdiwala S, Tyroler HA. High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies. *Circulation*. 1989;79:8–15.
43. Lee CD, Folsom AR, Nieto FJ, Chambless LE, Shahar E, Wolfe DA. White blood cell count and incidence of coronary heart disease and ischemic stroke, and mortality from cardiovascular disease in African-American and white men and women: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 2001;154:758–764.
44. Kaprio J. Twins and the mystery of missing heritability: the contribution of gene-environment interactions. *J Intern Med*. 2012;272:440–448.
45. Bentley AR, Sung YJ, Brown MR, Winkler TW, Kraja AT, Ntalla I, Schwander K, Chasman DI, Lim E, Deng X, et al. Multi-ancestry genome-wide gene-smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nat Genet*. 2019;51:636–648.
46. Dahl A, Nguyen K, Cai N, Gandal MJ, Flint J, Zaitlen N. A robust method uncovers significant context-specific heritability in diverse complex traits. *AJHG*. 2020;106:71-91.
47. Arena R, Ozemek C, Laddu D, Campbell T, Rouleau CR, Standley R, Bond S, Abril EP, Hills AP, Lavie CJ. Applying precision medicine to healthy living for the prevention and treatment of cardiovascular disease. *Curr Probl Cardiol*. 2018;43:448–483.
48. Leopold JA, Loscalzo J. Emerging role of precision medicine in cardiovascular disease. *Circ Res*. 2018;122:1302–1315.

Supplemental Material

Data S1. Extended description of statistical methods

Here we provide a description of the statistical models used in the study that is more formal than the description in the main text.

Multivariate Reaction Norm Model

We used a novel whole-genome modelling framework, Multivariate Reaction Norm Models (MRNMs), to detect G-C and R-C interactions. MRNM is an extension of bivariate linear mixed models. In the simplest form of a bivariate linear mixed model, the main trait, *y*, and the covariate, *c*, for individual *i*, after adjusting for their respective fixed effects, μ_y and μ_c, are simultaneously expressed as

$$\begin{pmatrix} y_i - \mu_y \\ c_i - \mu_c \end{pmatrix} = \begin{pmatrix} g_i \\ \beta_i \end{pmatrix} + \begin{pmatrix} e_i \\ \varepsilon_i \end{pmatrix}, \qquad \text{Equation 1}$$

Where g_i ~ N(0, σ_g^2) and β_i ~ N(0, σ_β^2) are genetic effects, which are aggregates of random effects of genome-wide SNPs on the main trait and on the covariate, respectively; e_i ~ N(0, σ_e^2) and ε_i ~ N(0, σ_ε^2) are residual effects, and both g_i and β_i are independent from e_i and ε_i.

MRNM extends Equation 1 by decomposing the random effects of the main traits into main effects and effects modulated by the covariate, which can be written as

$$\begin{pmatrix} y_i - \mu_y \\ c_i - \mu_c \end{pmatrix} = \begin{pmatrix} g_i \\ \beta_i \end{pmatrix} + \begin{pmatrix} e_i \\ \varepsilon_i \end{pmatrix} = \begin{pmatrix} \alpha_{0i} + c_i \cdot \alpha_{1i} \\ \beta_i \end{pmatrix} + \begin{pmatrix} \tau_{0i} + c_i \cdot \tau_{1i} \\ \varepsilon_i \end{pmatrix} \qquad \text{Equation 2}$$

where g_i breaks into $\alpha_{0i} + c_i \cdot \alpha_{1i}$, e_i into $\tau_{0i} + c_i \cdot \tau_{1i}$, α_{0i} ~ N(0, $\sigma_{\alpha 0}^2$), α_{1i} ~ N(0, $\sigma_{\alpha 1}^2$), τ_{0i} ~ N(0, $\sigma_{\tau 0}^2$) and τ_{1i} ~ N(0, $\sigma_{\tau 1}^2$). We use c_i to denote the covariate for individual *i*, α_{0i} for the main genetic effect on the main trait y_i , τ_{0i} for the residual effect, α_{1i} for the genotype-covariate interaction effect, and τ_{1i} for the residual-covariate interaction effect.

As shown in both equations, variance of the main trait and of the covariate are partitioned into two general sources, one of genetics (i.e., σ_g^2 & σ_β^2) and one of non-genetics or residuals (i.e., σ_e^2 & σ_ε^2). By modelling the main trait and the covariate simultaneously, the covariance between the main trait and the covariate, in forms of $cov\,(g_i\,, \beta_i)$ and $cov\,(e_i\,, \varepsilon_i)$, is accounted for in a MRNM. This is important given that the covariance between the main trait and covariate can sometimes be nontrivial and would have been neglected in univariate random regression models. More importantly though, the g_i and e_i terms are expanded in terms of c_i in Equation 2, which offers opportunities to model the genetic and residual variances of the main trait as a function of the covariate. With this expansion, it is immediately clear that genetic variance σ_g^2 breaks into var$(\alpha_0 + c \cdot \alpha_1)$ and residual variance σ_e^2 into var$(\tau_0 + c \cdot \tau_1)$, both of which vary with respect to the covariate. As such, MRNMs can estimate and detect genetic and residual variance heterogeneity due to the chosen covariate. A G-C interaction that underlies genetic variance heterogeneity is indicated by significant $\sigma_{\alpha 1}^2$,

and a R-C interaction that underlies residual variance heterogeneity is indicated by significant $\sigma^2_{\tau 1}$.

The MRNM shown in Equation 2 is referred to as the **full** model, which assumes and detects both genetic and residual variance heterogeneity with respect to the covariate. The full model can be simplified into other three major forms. Specifically, by setting both var($c \cdot \alpha_1$) and var($c \cdot \tau_1$) to 0, the **null** model assumes no heterogeneity in either the genetic or the residual variance of the main trait with respect to the covariate. By setting var($c \cdot \tau_1$) to 0, the **G-C** model assumes no R-C interaction and estimates the extent of genetic heterogeneity with respect to the covariate. Finally, by setting var($c \cdot \alpha_1$) to 0, the **R-C** model assumes no G-C interaction and estimates the extent of residual heterogeneity with respect to the covariate.

A detailed description of the variance-covariance structure assumed by MRNMs is provided below. Following Equation 1, the main trait, *y*, and the covariate, *c*, for individual *i*, after adjusting for their respective fixed effects, μ_y and μ_c, are simultaneously expressed as

$$\begin{pmatrix} y_i - \mu_y \\ c_i - \mu_c \end{pmatrix} = \begin{pmatrix} y_i^* \\ c_i^* \end{pmatrix} = \begin{pmatrix} g_i \\ \beta_i \end{pmatrix} + \begin{pmatrix} e_i \\ \varepsilon_i \end{pmatrix}$$

The variance-covariance matrix for *N* realizations of $\begin{pmatrix} y^* \\ c^* \end{pmatrix}$ can be expressed as

$$var\begin{pmatrix} y^* \\ c^* \end{pmatrix} = \begin{bmatrix} \mathbf{Z}_1\mathbf{A}\sigma^2_{g_1}\mathbf{Z}_1' + \mathbf{Z}_1\mathbf{I}\sigma^2_{e_1}\mathbf{Z}_1' & \cdots & \mathbf{Z}_1\mathbf{A}\sigma_{g_{1,N}}\mathbf{Z}_N' + \mathbf{Z}_1\mathbf{I}\sigma_{e_{1,N}}\mathbf{Z}_N' & \mathbf{Z}_1\mathbf{A}\sigma_{g_{1,\beta}}\mathbf{Z}_c' + \mathbf{Z}_1\mathbf{I}\sigma_{e_{1,\varepsilon}}\mathbf{Z}_c' \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{Z}_N\mathbf{A}\sigma_{g_{1,N}}\mathbf{Z}_1' + \mathbf{Z}_N\mathbf{I}\sigma_{e_{1,N}}\mathbf{Z}_1' & \cdots & \mathbf{Z}_N\mathbf{A}\sigma^2_{g_N}\mathbf{Z}_N' + \mathbf{Z}_N\mathbf{I}\sigma^2_{e_N}\mathbf{Z}_N' & \mathbf{Z}_N\mathbf{A}\sigma_{g_N,\beta}\mathbf{Z}_c' + \mathbf{Z}_N\mathbf{I}\sigma_{e_N,\varepsilon}\mathbf{Z}_c' \\ \mathbf{Z}_c\mathbf{A}\sigma_{g_1,\beta}\mathbf{Z}_1' + \mathbf{Z}_c\mathbf{I}\sigma_{e_1,\varepsilon}\mathbf{Z}_1' & \cdots & \mathbf{Z}_c\mathbf{A}\sigma_{g_N,\beta}\mathbf{Z}_N' + \mathbf{Z}_c\mathbf{I}\sigma_{e_N,\varepsilon}\mathbf{Z}_N' & \mathbf{Z}_c\mathbf{A}\sigma^2_{\beta}\mathbf{Z}_c' + \mathbf{Z}_c\mathbf{I}\sigma^2_{\varepsilon}\mathbf{Z}_c' \end{bmatrix}$$

Where $\mathbf{I}$ is an *N* x *N* identity matrix, **A** is the *N* x *N* genomic relationship matrix based on genome-wide SNP information, $\mathbf{Z}_i$ is the incident matrix for g_i for i = 1, 2 …*N*, and $\mathbf{Z}_c$ is the incident matrix for $\boldsymbol{c}$.

Prior to model fitting, we attempted to simplify the general MRNMs outlined above by reducing the number of free parameters for estimation. We estimated heritability of each lifestyle covariate in the ARIC dataset via univariate Genomic Restricted Maximum Likelihood (GREML), and found that all estimates were close to zero. Daily potassium intake was the only covariate with an estimate marginally different from zero (h^2 = 0.08 ± 0.04). Subsequently, we simplified MRNMs by setting σ^2_β (i.e., genetic variance of the covariate) and its associated covariance terms, i.e., $cov\,(\alpha_0\,, \beta)$ and $cov\,(\alpha_1\,, \beta)$, to 0. Unless specified otherwise, all MRNMs fitted to ARIC data in this paper are simplified MRNMs.

For each pair of main trait and covariate, the null and full models were fitted and compared using a likelihood ratio test. For the simplified MRNMs, the test statistic, i.e., -2 log likelihood ratio, is assumed to have a chi-square distribution with five degrees of freedom. The alpha level was set at 0.05. A significant p-value indicates the full

model has a better fit than the null, hence the presence of a G-C, R-C interaction, or both. Since the full model does not separate G-C and R-C interactions, we considered a model comparison strategy to separate the two. However, we show in Data S3 that this strategy can suffer from weak statistical power and biased estimation, which makes it an overall inferior method to the null versus full model comparison method for detecting G-C and R-C interactions. Therefore, our results are based on the latter. All model fitting for this paper was performed using MTG2.

UKBB Validation

To validate significant results found in the ARIC dataset, we repeated analyses using the UKBB for variables where the two datasets overlap. Since the UKBB has a larger sample size, hence greater statistical power, we explicitly estimated the genetic variance of the covariate when fitting a MRNM, rather than fixing this parameter at zero as for the ARIC dataset. Subsequently, the degree of freedom used for the likelihood ratio test that compares the full model with the null model was seven as opposed to five. Same as for the ARIC dataset, we estimated heritability of each UKBB trait using two URNMS (i.e., null and interaction models) and the inclusion of a covariate was based on MRNM results.

Heritability Models

We considered the consequence of neglecting G-C and R-C interactions on heritability estimates. Specifically, we estimated heritability of each trait using two models, one that includes no interaction term at all, i.e., null model (also known as GREML), and the other that includes one or more interaction terms, i.e., interaction model, and compared estimates of the two models. To reduce computational burden, we used univariate reaction norm models (URNMs), as opposed to MRNMs. The null model in the univariate framework is essentially Equation 1 without the part that involves the covariate, c_i. Using the same notation as Equation 1, the main trait for individual i, in a URNM can be written as:

Null model: $$y_i - \mu_y = g_i + e_i$$

The interaction model in the univariate framework expands g_i and e_i as functions of m_1 and m_2 covariates, respectively, where $m_1 + m_2 \geq 1$. Using j to index covariate, the main trait for individual i in a URNM with interaction terms can be written as:

Interaction model: $$y_i - \mu_y = \alpha_{0i} + \sum_{j=1}^{m_1} c_{ij}\alpha_{ij} + \tau_{0i} + \sum_{k=1}^{m_2} c_{ik}\tau_{ik}$$

Data S2. Simulation studies

Simulation Settings

To facilitate data interpretation, we simulated phenotypic data with and without G-C and/or R-C interactions and assessed, using simulated data, whether MRNMs can produce unbiased parameter estimates, type I error rate and power of detecting G-C and R-C interactions. We purposely chose two sets of model parameter configurations that varied primarily in effect size for heritability, G-C and R-C interactions. One setting had large effect sizes, referred to as the 'large-effects setting', with a heritability of 0.5 for both the main trait and the covariate and both $\sigma^2_{\alpha 1}$ and $\sigma^2_{\tau 1}$, which are indicative of G-C and R-C interactions, were set at 0.5. In contrast, the other setting had smaller effect sizes, referred to as the 'small-effects setting', with a heritability of 0.15 for the main trait, 0 for the covariate, and both $\sigma^2_{\alpha 1}$ and $\sigma^2_{\tau 1}$ were set at 0.05. It is noted that the small-effects setting resembled more closely parameter estimates from real data analyses than the large-effects setting. Thus, results of the former setting would be more informative about how well our models and the likelihood test for model comparisons perform for analysis of real data.

Each parameter setting covered four scenarios—no G-C and R-C interactions (or the null), R-C interaction only, G-C interaction only, and both R-C and G-C interactions—where the true data generating models were the four models described above. Under each scenario, we simulated 100 replicates of phenotypic data (n = 7,513) of a main trait and a covariate, each based on 10,000 randomly chosen causal variants from the ARIC genotype data (see Table S1 for an overview). For every replicate, we fitted the full and null models and compared the fit of the two models using the abovementioned likelihood ratio test. For every scenario, we computed the proportion of replicates, out of 100, for which the full model has a better fit than the null. This proportion takes on different interpretations depending on the simulation scenario. It is an estimate of type I error rate when the true model is the null, whereas it is an estimate of statistical power in scenarios where the true model is other than the null.

It is important to note that all simulating models above assume normally distributed random effects (e.g., genetic and residual effects). In effect, for any given covariate value, the main trait follows a normal distribution. This normality assumption however, is likely violated for many traits of the ARIC and UKBB datasets, which are characterised by substantially larger kurtosis and skewness than would be expected from data simulated under normality (Figure S1). Therefore, in addition to the large and small effects settings described above, we also simulated data with non-normal residuals drawn from Gamma distributions. We purposefully chose two sets of shape and scale parameters of Gamma distributions to represent large and small deviations from normality. For each of the non-normal settings, we had two scenarios: no G-C and R-C interaction (i.e., the null model is true) and G-C and R-C interactions (i.e., the full model is true), each with 100 replicates. We fitted the null and full models, which by definition all assume normality of random effects, to each replicate and subsequently assessed our model comparison method, in terms of type I error, power,

and parameter estimates. In the event of an inflated type I error rate, we applied a rank-based inverse normal transformation to the simulated data and refitted the models. We then assessed the effectiveness of the transformation on reducing false positive findings and its potential consequences on statistical power of the model comparison method and model parameter estimates.

It is important to emphasize that our interaction models do not assume absolute normality of phenotype data, rather their conditional normality on the covariate. Thus, unless the true underlying model is the null, when phenotypic observations are collapsed across covariate values, the distribution of the collapsed data is not necessarily normal. In fact, in the presence of genuine G-C and/or R-C interactions, even when the model normality assumption is met, the simulated phenotype can have larger skewness and kurtosis than data simulated under the null model (see Figure S1). Thus, deviations from normality of a given set of phenotype data could arise from genuine G-C or R-C interaction. If they are mistaken as signs of violation of the model normality assumption, what would be the consequences of applying a rank-based inverse normal transformation for type I error rate, statistical power and model estimates? To answer these questions, we also applied the transformation to phenotype data simulated under normality (i.e., large & small effect parameter settings) and assessed its impact on type I error rate, statistical power and model estimates.

Simulation Results

When the model assumption of normality was met, the estimated type I error rate of the null versus full model comparison method was not inflated (0.04; see Table S2). Small and large phenotypic deviations from the normality inflated the type I error rate to 0.2 and 0.65, respectively. However, after a rank-based inverse normal transformation (RINT) of phenotypic data, the type I error rate was approximately controlled (0.05 and 0.07 for large and small phenotypic deviations from normality, respectively; Table S2), indicating that an RINT can effectively reduce false positive findings in face of violations of the normality assumption held by the MRNMs.

The statistical power of the null versus full model comparison was estimated using data simulated under scenarios other than the null, i.e., G-C only, R-C only and both G-C and R-C interactions. We found that whether the normality assumption is met or not, the proportion of replicates for which the full model had a better fit than the null was at least 0.88 (Table S2), giving an estimated power above 88%. Applying an RINT did not affect the power in any scenario.

For each simulation scenario, we compared parameter estimates from the full model with their corresponding true values. Figure S2 shows sampling distributions of full-model parameter estimates based on 100 replicates for both large and small effects settings (in terms of heritability, G-C and R-C interactions; Table S1) when the model assumption of normality was met, and it indicates that the full model produced unbiased estimates of model parameters under all simulation scenarios. This observation holds even when the normality assumption was violated (Figures S3 & S4). In contrast, after applying an RINT, full-model estimates were biased for some model parameters (Figures S3 & S4).

In summary, our simulation results indicate that when the model assumption of normality is met, the likelihood ratio test that compares the full model with the null can detect G-C and/or R-C interaction at an acceptable type I error rate with a reasonable level of power. When the normality assumption is violated, however, type I error rate would be inflated, in which case an RINT of the phenotype data is an effective remedy without compromising statistical power. In situations where the normality assumption is not violated, a rank-based inverse normal transformation of the phenotype data would not adversely affect type I error rate and statistical power. In terms of parameter estimates, full-model estimates of heritability, G-C and R-C interactions are unbiased, regardless of whether or not the normality assumption is violated. Full-model estimates would however become biased after an RINT. Therefore, for analysis of real data, if the model assumption of normality is in doubt, rank-based inverse transformation should be applied to control type I error rate; and once a significant finding is declared, full model estimates of parameters from data without the transformation should be reported and interpreted.

Data S3. Alternative model comparison strategy

Throughout the text we relied on the null versus full model comparison for detecting G-C and R-C interactions. An alternative and seemingly more logical strategy would be to derive the best model via model comparisons that involve reduced models, i.e., G-C only and R-C only models, in addition to the null and full models. Here we elaborate on this alternative strategy; and using results from simulations, we further show issues associated with this strategy. We conclude that the null versus full model comparison is a superior method, hence a logical choice for detecting G-C and R-C interactions.

This alternative strategy uses result patterns from four model comparisons to conclude the best model out of five candidates (Table S3). Candidates considered were the null model, G-C interaction only model, R-C interaction only model, full model, and G-C or R-C interaction models. The last candidate is more of a situation than a model, where the model comparison method does not distinguish between G-C and R-C models, that is, model selection is inconclusive. It occurs when both G-C model and R-C model show a better fit than the null but a worse fit than the full, that is, G-C and R-C models are equally likely. Upon concluding the best model, the source of variance heterogeneity is immediately implied. For example, genetic variance heterogeneity (i.e., a G-C interaction) is declared, when either the G-C model or the full model is the best. In contrast, residual variance heterogeneity (i.e., a R-C interaction) is declared, when either the R-C model or the full model is the best.

We applied this model selection strategy to simulated data from large- and small-effects settings and evaluated how well the strategy can recover the true simulating model. The results are summarised in Table S4. The method correctly identified the null model for at least 97% of the simulated replicates, giving an estimated type I error rate of 0.03, which is well under the 0.05 target. However, statistical power—estimated by the proportion of replicates for which a true model other than the null is correctly identified—varied largely depending on effect size. For the large-effects setting, a true model other than the null was correctly identified for at least 91% of replicates, hence an estimated power of 0.91 and above. In contrast, for the small-effects setting, the estimated power was 0.04 at worst and 0.11 at best. This does not mean though, the model comparison method could not detect G-C and R-C interactions that are small in magnitude. Rather, for over 75% of replicates under this setting, the likelihood ratio test results were such that G-C and R-C models fit data equally well (see last column of Table S4). In short, either when there are no genuine G-C and R-C interactions or when genuine G-C and R-C interactions are large in magnitude, the likelihood-ratio-based method can discern the true underlying model at a high accuracy (>0.9). However, when genuine G-C and R-C interactions are small, it is unlikely that the method will uncover the true model.

For each simulation scenario, we also compared parameter estimates from all four fitted models with their corresponding true values and noted that results are similar for the two parameter settings that vary in effect sizes. Figure S5 shows results for the small-effects setting, which hold for the large-effects setting. When the true underlying model was the null, regardless of which model was fitted, all parameter estimates were

unbiased. In scenarios where the true model was other than the null, fitting the correct model produced unbiased estimates for all parameters. However, in these scenarios, fitting a wrong model—that is, a model other than the true—could produce biased estimates for some parameters. For example, fitting the null model to data with genuine G-C or/and R-C interactions produced larger estimates of the residual variance, i.e., $\delta^2_{\tau 0}$, than its true value, by an amount similar to the set value of $\delta^2_{\alpha 1}$ or/and $\delta^2_{\tau 1}$. When fitting the G-C model to data with R-C interaction but no G-C interaction, estimates of $\delta^2_{\alpha 1}$ were larger than the true, i.e., 0, by an amount similar to the set value of $\delta^2_{\tau 1}$. Likewise, when fitting the R-C model to data with G-C interaction but no R-C interaction, estimates of $\delta^2_{\tau 1}$ deviated from the true by an amount similar to the set value of $\delta^2_{\alpha 1}$. However, fitting the full model, even when it was the wrong model, produced unbiased estimates for all parameters. Thus, our simulation results indicate that model misspecification can result in biased estimates for some parameters depending on the simulation scenario, with the only exception of fitting the full model, which provides unbiased estimates for all parameters in all scenarios.

In summary, the model selection strategy has a type I error rate under 0.05, but its power of recovering the true model is very low when effect sizes are small. Consequently, this strategy can result in an alarmingly elevated chance of concluding a wrong model, i.e., model misspecification, which could produce biased estimates for some model parameters. Therefore, for analysis of real data, where the true underlying model is unknown and effects sizes are likely small, this model comparison strategy is not useful to select the best model for identifying source of variance heterogeneity. In contrast, we showed in the main text that the null versus full model comparison method has an acceptable type I error rate and reasonable power when effect sizes are small. Even if the full model is not true, model estimates are not biased, which can be interpreted subsequently. Hence, the null versus full model comparison is a superior method to the alternative model comparison strategy.

Table S1. True parameter values of four simulation models under four settings.

A. Small effects under Normality

Parameter	*No G-C & R-C*	*G-C & R-C*	*G-C only*	*R-C only*
var(α_0)	0.15	0.15	0.15	0.15
var(α_1)	0	0.05	0.05	0
var(τ_0)	0.85	0.85	0.85	0.85
var(τ_1)	0	0.05	0	0.05
var(β)	0	0	0	0
var(ε)	1	1	1	1
cov(α_0, α_1)	0	0	0	0
cov(τ_0, τ_1)	0	0	0	0
cov(α_0, β)	0	0	0	0
cov(α_1, β)	0	0	0	0
cov(τ_0, ε)	0	0	0	0
cov(τ_1, ε)	0	0	0	0

B. Large effects under Normality

Parameter	*No G-C & R-C*	*G-C & R-C*	*G-C only*	*R-C only*
var(α_0)	0.5	0.5	0.5	0.5
var(α_1)	0	0.5	0.5	0
var(τ_0)	0.5	0.5	0.5	0.5
var(τ_1)	0	0.5	0	0.5
var(β)	0.5	0.5	0.5	0.5
var(ε)	0.5	0.5	0.5	0.5
cov(α_0, α_1)	0	0.05	0.05	0
cov(τ_0, τ_1)	0	0.05	0	0.05
cov(α_0, β)	0	0	0	0
cov(α_1, β)	0	0	0	0
cov(τ_0, ε)	0	0	0	0
cov(τ_1, ε)	0	0	0	0

C. Small Deviation from Normality

Parameter	*No G-C & R-C*	*G-C & R-C*	*G-C only*	*R-C only*
var(α_0)	0.15	0.15	0.15	0.15
var(α_1)	0	0.05	0.05	0
var(τ_0)	0.85	0.85	0.85	0.85
var(τ_1)	0	0.05	0	0.05
var(β)	0	0	0	0
var(ε)	1	1	1	1
cov(α_0, α_1)	0	0	0	0
cov(τ_0, τ_1)	0	0	0	0
cov(α_0, β)	0	0	0	0
cov(α_1, β)	0	0	0	0
cov(τ_0, ε)	0	0	0	0
cov(τ_1, ε)	0	0	0	0
k_0	2	2	2	2
θ_0	0.65	0.65	0.65	0.65
k_1	-	2	-	2
θ_1	-	0.16	-	0.16

D. Large Deviation from Normality

Parameter	*No G-C & R-C*	*G-C & R-C*	*G-C only*	*R-C only*
var(α_0)	0.15	0.15	0.15	0.15
var(α_1)	0	0.05	0.05	0
var(τ_0)	0.85	0.85	0.85	0.85
var(τ_1)	0	0.05	0	0.05
var(β)	0	0	0	0
var(ε)	1	1	1	1
cov(α_0, α_1)	0	0	0	0
cov(τ_0, τ_1)	0	0	0	0
cov(α_0, β)	0	0	0	0
cov(α_1, β)	0	0	0	0
cov(τ_0, ε)	0	0	0	0
cov(τ_1, ε)	0	0	0	0
k_0	0.25	0.25	0.25	0.25
θ_0	1.84	1.84	1.84	1.84
k_1	-	0.25	-	0.25
θ_1	-	0.45	-	0.45

The top two settings are under the assumption that all random effects of the multivariate reaction normal models (MRNMs) for simulation are drawn from normal distributions. This assumption is relaxed for the bottom two settings, where residual effects, τ_0 and τ_1, are drawn from Gamma(k_0, θ_0) and Gamma(k_1, θ_1) with mean centred at zero, respectively. Each setting comprises four simulation models, which from the left to right are the null, full, G-C and R-C models.

Table S2. Proportion of simulated replicates for which the full model had a better fit than the null under different simulation scenarios.

Simulation model	Simulation under Normality				Simulation under Non-Normality				Interpretation
	large effects		small effects		large deviation		small deviation		
	*no RINT**	*RINT*	*no RINT*	*RINT*	*no RINT*	*RINT*	*no RINT*	*RINT*	
No G-C & R-C	0.04	0.04	0.04	0.05	0.65	0.05	0.2	0.07	type I error
G-C only	1	1	0.93	0.94	0.81	1	0.84	0.99	power
R-C only	1	1	0.88	0.87	0.84	1	0.87	0.99	power
G-C & R-C	1	1	1	1	1	1	1	1	power

*RINT = Rank-based Inverse Normal Transformation

Data of a main trait and a covariate were simulated using four models (1st column) under normality with large and small effects (in terms of heritability, Genotype-Covariate (G-C) and Residual-Covariate (R-C) interactions) and under non-normality that resulted in large and small phenotypic deviations of the main trait from normality. Each simulation was repeated 100 times, resulting in 100 replicates of simulated data under each setting. For each replicate, the full model, which allows G-C and R-C interactions, and the null model, which assumes no G-C and R-C interactions, were fitted then compared using a likelihood ratio test. The model comparison was repeated after a rank-based inverse normal transformation was applied to the simulated data.

Table S3. Overview of Model Selection Strategy.

Model Comparison	Candidate Model				
	Null	G-C	R-C	Full	G-C/R-C
Null vs. R-C	×		✓	✓	✓
Null vs. G-C	×	✓		✓	✓
R-C vs. Full		✓	×	✓	×
G-C vs. Full		×	✓	✓	×

Each column shows the model comparison result pattern required to conclude a given candidate model is the best. Result patterns are mutually exclusive across the five candidates. A cross indicates a non-significant p-value for a model comparison (i.e., the simpler model is better), whereas a tick indicates a significant p-value (i.e., the simpler model is worse). Comparisons that are not necessary for model selection are left as blanks. Note the pattern in the last column does not distinguish between Genotype-Covariate (G-C) and Residual-Covariate (R-C) interactions, in which case model selection is inconclusive.

Table S4. Proportions of simulated replicates for which a given candidate model is chosen as the best under different simulation scenarios.

	Simulation Scenario	Candidate Model				
		Null	G-C	R-C	Full	G-C/R-C
large-effects	No G-C & R-C	**0.98**	0	0	0	0.02
	G-C only	0	**0.93**	0	0.07	0
	R-C only	0	0	**0.96**	0.04	0
	G-C & R-C	0	0.03	0.06	**0.91**	0
small-effects	No G-C & R-C	**0.97**	0.01	0.02	0	0
	G-C only	0.06	**0.11**	0.03	0.04	0.76
	R-C only	0.1	0.01	**0.09**	0.05	0.75
	G-C & R-C	0	0.05	0.14	**0.04**	0.77

Data were simulated under normality with large and small effects in terms of heritability, Genotype-Covariate (G-C) and Residual-Covariate (R-C) interactions. Each scenario had 100 replicates of a main trait and a covariate. For each replicate, four models were fitted and compared to select the best fitting one (see Table S3).

Table S5. Variance and covariance estimates from the full model for the 34 signals emerged from the atherosclerosis risk in communities study.

Main Trait	Lifestlye Covariate	var(α_0)	var(α_1)	cov(α_0,α_1)	var(τ_0)	var(τ_1)	cov(τ_0,τ_1)
Fibrinogen	Cigarette years of smoking	4.0e+02(1.5e+02)	-3.6e+02(1.5e+02)	-7.9e+01(1.1e+02)	3.1e+03(1.6e+02)	2.9e+02(1.6e+02)	3.6e+02(1.2e+02)
Fibrinogen	Physical activity:sports domain	4.2e+02(1.5e+02)	-2.5e+02(1.3e+02)	1.5e+01(1.0e+02)	2.9e+03(1.6e+02)	3.6e+02(1.4e+02)	-2.7e+02(1.1e+02)
Factor VII	Physical activity:sports domain	1.0e+02(3.6e+01)	2.6e+00(3.1e+01)	-1.5e+01(2.4e+01)	6.5e+02(3.8e+01)	7.2e+00(3.2e+01)	-4.9e+01(2.5e+01)
BMI	Alcohol intake (g/week)	2.9e+00(9.0e-01)	1.8e-01(6.9e-01)	-1.1e+00(5.7e-01)	1.7e+01(9.3e-01)	5.7e-01(6.9e-01)	-1.3e+00(6.0e-01)
BMI	Physical activity:leisure domain	3.2e+00(9.2e-01)	-6.7e-02(8.3e-01)	3.1e-01(6.3e-01)	1.7e+01(9.6e-01)	5.9e-01(8.7e-01)	-1.3e+00(6.4e-01)
BMI	Physical activity:sports domain	3.1e+00(9.1e-01)	1.1e-01(6.6e-01)	2.3e-02(5.8e-01)	1.7e+01(9.6e-01)	4.2e-01(7.1e-01)	-2.5e+00(6.1e-01)
BMI	Keys score	3.2e+00(9.2e-01)	-4.0e-01(8.0e-01)	-3.3e-01(6.2e-01)	1.7e+01(9.6e-01)	5.5e-01(8.3e-01)	1.5e+00(6.3e-01)
BMI	Saturated fat intake (g/day)	3.1e+00(9.2e-01)	4.3e-01(8.7e-01)	-2.5e-01(6.5e-01)	1.7e+01(9.6e-01)	-5.8e-01(9.1e-01)	1.2e+00(6.6e-01)
BMI	Energy from saturated fat (%kcal/day)	3.1e+00(9.2e-01)	-1.2e+00(8.1e-01)	-5.2e-01(6.2e-01)	1.7e+01(9.6e-01)	1.5e+00(8.5e-01)	1.6e+00(6.4e-01)
BMI	Energy from total fat intake (%kcal/day)	3.1e+00(9.2e-01)	-1.2e+00(8.2e-01)	-3.4e-01(6.3e-01)	1.7e+01(9.6e-01)	1.5e+00(8.8e-01)	1.3e+00(6.4e-01)
Waist-to-Hip Ratio	Cigarette years of smoking	3.0e-04(2.0e-04)	0.0e+00(2.0e-04)	-1.0e-04(1.0e-04)	3.7e-03(2.0e-04)	0.0e+00(2.0e-04)	-1.0e-04(1.0e-04)
Waist-to-Hip Ratio	Physical activity:sports domain	3.0e-04(2.0e-04)	-2.0e-04(1.0e-04)	0.0e+00(1.0e-04)	3.8e-03(2.0e-04)	3.0e-04(2.0e-04)	-2.0e-04(1.0e-04)
Waist-to-Hip Ratio	Total energy intake (kcal/day)	3.0e-04(2.0e-04)	-1.0e-04(1.0e-04)	1.0e-04(1.0e-04)	3.9e-03(2.0e-04)	0.0e+00(2.0e-04)	-3.0e-04(1.0e-04)
Waist-to-Hip Ratio	Energy from protein intake (%kcal/day)	3.0e-04(2.0e-04)	-1.0e-04(2.0e-04)	-1.0e-04(1.0e-04)	3.8e-03(2.0e-04)	0.0e+00(2.0e-04)	2.0e-04(1.0e-04)
Pulse Pressure	Cigarette years of smoking	8.0e+00(5.1e+00)	-1.1e+00(5.5e+00)	-2.7e+00(3.8e+00)	1.0e+02(5.5e+00)	6.3e+00(5.8e+00)	4.6e+00(4.1e+00)
Diastolic Blood Pressure	Cigarette years of smoking	7.4e+00(3.2e+00)	-7.3e+00(3.0e+00)	-2.5e+00(2.2e+00)	6.4e+01(3.4e+00)	8.1e+00(3.2e+00)	4.4e+00(2.4e+00)
Heart Rate	Cigarette years of smoking	1.3e+01(4.1e+00)	-2.1e+00(4.3e+00)	1.9e+00(3.0e+00)	7.1e+01(4.3e+00)	6.8e+00(4.6e+00)	-1.6e+00(3.1e+00)
Heart Rate	Physical activity:leisure domain	1.3e+01(4.1e+00)	-4.4e+00(3.4e+00)	4.9e+00(2.7e+00)	7.4e+01(4.3e+00)	5.9e+00(3.7e+00)	-8.8e+00(2.8e+00)
Heart Rate	Physical activity:sports domain	1.2e+01(4.0e+00)	1.1e+00(3.2e+00)	3.4e+00(2.6e+00)	7.7e+01(4.3e+00)	-4.6e-01(3.4e+00)	-6.7e+00(2.7e+00)
HDL2 Cholesterol	Cigarette years of smoking	-1.0e-04(1.8e-03)	-3.0e-04(1.3e-03)	5.0e-04(1.1e-03)	4.0e-02(1.9e-03)	4.0e-04(1.4e-03)	-3.1e-03(1.1e-03)
HDL2 Cholesterol	Physical activity:leisure domain	-6.0e-04(1.8e-03)	-1.2e-03(1.7e-03)	2.0e-04(1.2e-03)	4.0e-02(1.9e-03)	1.7e-03(1.8e-03)	3.0e-03(1.3e-03)
HDL2 Cholesterol	Carbohydrate intake (g/day)	-5.0e-04(1.8e-03)	-1.1e-03(1.4e-03)	-4.0e-04(1.1e-03)	4.0e-02(1.9e-03)	1.2e-03(1.4e-03)	-1.9e-03(1.1e-03)
HDL2 Cholesterol	Total energy intake (kcal/day)	-4.0e-04(1.8e-03)	-1.5e-03(1.3e-03)	-6.0e-04(1.1e-03)	4.1e-02(1.9e-03)	6.0e-04(1.4e-03)	-9.0e-04(1.2e-03)
HDL2 Cholesterol	Energy from protein intake (%kcal/day)	-4.0e-04(1.8e-03)	1.6e-03(1.6e-03)	-4.0e-04(1.2e-03)	4.1e-02(1.9e-03)	-2.2e-03(1.6e-03)	2.4e-03(1.3e-03)
HDL3 Cholesterol	Alcohol intake (g/week)	5.9e-03(2.7e-03)	1.0e-03(3.3e-03)	-3.9e-03(2.1e-03)	5.5e-02(2.9e-03)	-7.0e-04(3.3e-03)	5.9e-03(2.3e-03)
HDL3 Cholesterol	Physical activity:sports domain	6.2e-03(2.7e-03)	5.0e-03(2.8e-03)	1.4e-03(2.0e-03)	5.1e-02(2.9e-03)	-1.8e-03(2.9e-03)	-2.4e-03(2.0e-03)
HDL Cholesterol	Physical activity:sports domain	1.4e-02(6.1e-03)	1.4e-02(6.3e-03)	4.8e-03(4.3e-03)	1.2e-01(6.6e-03)	-7.8e-03(6.4e-03)	-7.0e-03(4.4e-03)
HDL Cholesterol	Monounsaturated fatty acid intake (g/day)	1.3e-02(6.2e-03)	9.0e-04(5.2e-03)	-2.0e-03(4.0e-03)	1.2e-01(6.5e-03)	-9.0e-04(5.2e-03)	-4.4e-03(4.2e-03)
HDL Cholesterol	Energy from protein intake (%kcal/day)	1.3e-02(6.2e-03)	-4.7e-03(5.9e-03)	1.0e-03(4.4e-03)	1.2e-01(6.6e-03)	5.8e-03(6.1e-03)	4.3e-03(4.5e-03)
Apolipoprotein AI	Polyunsaturated fatty acid intake (g/day)	7.3e+03(3.4e+03)	-3.4e+03(2.9e+03)	3.7e+03(2.1e+03)	6.8e+04(3.5e+03)	3.8e+03(2.9e+03)	-7.4e+03(2.3e+03)
White Blood Cell Count	Cigarette years of smoking	4.1e-01(1.2e-01)	-6.6e-01(4.6e-02)	-3.0e-01(8.0e-02)	3.0e+00(1.4e-01)	5.8e-01(7.4e-02)	8.6e-01(9.2e-02)
White Blood Cell Count	Physical activity:leisure domain	5.7e-01(1.3e-01)	1.1e-01(1.3e-01)	-7.1e-02(9.4e-02)	2.3e+00(1.4e-01)	1.2e-02(1.3e-01)	-1.7e-01(9.5e-02)
White Blood Cell Count	Physical activity:sports domain	5.5e-01(1.3e-01)	1.4e-01(1.3e-01)	-1.3e-01(9.5e-02)	2.2e+00(1.4e-01)	9.1e-02(1.4e-01)	-9.1e-02(9.7e-02)
White Blood Cell Count	Keys score	5.5e-01(1.3e-01)	3.9e-03(1.2e-01)	1.5e-01(9.3e-02)	2.4e+00(1.4e-01)	-1.0e-02(1.3e-01)	3.0e-03(9.3e-02)

All estimates are derived from analyses of data without a rank-based inverse normal transformation. Standard errors are in brackets. Other model parameters are omitted for simplicity. Signals replicated in the UK biobank are shaded. Note the UK biobank only has data available to validate 17 signals emerged from ARIC.

Table S6. Variance and covariance estimates from the full model for the UK Biobank dataset.

Main Trait	Lifestlye Covariate	var(α_0)	var(α_1)	cov(α_0,α_1)	var(τ_0)	var(τ_1)	cov(τ_0,τ_1)
BMI	Alcohol intake (glass & pint/week)	3.9e+00(2.0e-01)	4.4e-01(2.1e-01)	-1.4e-01(1.4e-01)	1.3e+01(2.1e-01)	3.2e-01(2.3e-01)	-4.4e-01(1.6e-01)
BMI	MET minutes/week for walking	4.0e+00(2.0e-01)	5.7e-02(1.7e-01)	-3.6e-01(1.3e-01)	1.2e+01(2.3e-01)	1.1e+00(1.9e-01)	-1.2e+00(1.5e-01)
BMI	MET minutes/week for moderate activity	3.9e+00(2.0e-01)	-1.8e-01(1.4e-01)	-1.9e-01(1.2e-01)	1.2e+01(2.2e-01)	1.4e+00(1.8e-01)	-1.7e+00(1.5e-01)
BMI	MET minutes/week for vigorous activity	4.0e+00(2.1e-01)	1.6e-01(5.7e-01)	-1.6e-01(5.7e-01)	1.3e+01(2.2e-01)	5.8e-01(5.8e-01)	-1.8e+00(5.7e-01)
BMI	Summed MET minutes/week for all activity	3.9e+00(3.7e-01)	-1.1e-01(1.9e-01)	-2.7e-01(2.4e-01)	1.3e+01(2.1e-01)	1.2e+00(1.9e-01)	-1.7e+00(2.2e-01)
BMI	estimated saturated fat intake	3.9e+00(3.7e-01)	4.7e-01(3.6e-01)	-9.6e-02(2.6e-01)	1.3e+01(3.8e-01)	-2.1e-01(3.7e-01)	1.9e-01(2.6e-01)
Diastolic Blood Pressure[1]	Pack years adult smoking as proportion of life span exposed to smoking	1.6e+01(1.2e+00)	2.2e+00(9.5e-01)	-1.8e+00(7.8e-01)	9.1e+01(1.3e+00)	-2.7e+00(9.0e-01)	3.2e+00(8.5e-01)
Pulse Pressure[1]	Pack years adult smoking as proportion of life span exposed to smoking	2.5e+01(2.0e+00)	-9.5e-01(1.8e+00)	2.1e+00(1.3e+00)	1.5e+02(2.2e+00)	2.4e-02(1.7e+00)	2.0e+00(1.5e+00)
Heart Rate	Pack years adult smoking as proportion of life span exposed to smoking	2.0e+01(1.5e+00)	6.0e-01(1.5e+00)	-6.7e-01(1.1e+00)	1.1e+02(1.8e+00)	2.1e+00(1.7e+00)	3.3e+00(1.2e+00)
Heart Rate	MET minutes/week for walking	2.0e+01(1.5e+00)	7.5e-01(1.3e+00)	2.0e-01(9.7e-01)	1.1e+02(1.8e+00)	1.5e+00(1.5e+00)	-3.8e+00(1.2e+00)
Heart Rate	Summed MET minutes/week for all activity	2.0e+01(1.5e+00)	-6.8e-01(1.3e+00)	-6.6e-02(1.0e+00)	1.1e+02(1.7e+00)	4.7e+00(1.5e+00)	-4.5e+00(1.2e+00)
Waist-to-Hip Ratio[2]	Pack years adult smoking as proportion of life span exposed to smoking	1.5e-01(1.1e-02)	-3.6e-03(1.0e-02)	-4.7e-03(7.7e-03)	8.5e-01(1.3e-02)	4.1e-02(1.2e-02)	3.4e-03(9.1e-03)
Waist-to-Hip Ratio	MET minutes/week for moderate activity	1.5e-01(1.1e-02)	5.6e-03(1.0e-02)	-1.2e-02(7.3e-03)	8.1e-01(1.3e-02)	2.9e-02(1.2e-02)	-2.3e-02(9.6e-03)
Waist-to-Hip Ratio	MET minutes/week for vigorous activity	1.5e-01(1.1e-02)	3.6e-03(1.0e-02)	-1.3e-02(7.5e-03)	8.3e-01(1.2e-02)	1.8e-02(1.1e-02)	-2.5e-02(9.4e-03)
Waist-to-Hip Ratio	Summed MET minutes/week for all activity	1.5e-01(1.1e-02)	9.5e-03(1.0e-02)	-1.1e-02(7.5e-03)	8.2e-01(1.3e-02)	1.9e-02(1.1e-02)	-1.9e-02(8.9e-03)
Waist-to-Hip Ratio	estimated total energy intake	1.8e-01(2.2e-02)	-2.0e-02(2.0e-02)	5.0e-03(1.5e-02)	8.1e-01(2.3e-02)	3.3e-02(2.1e-02)	-1.6e-02(1.5e-02)
White Blood Cell Count	Pack years adult smoking as proportion of life span exposed to smoking	5.2e-01(4.2e-02)	-2.7e-02(4.2e-02)	-2.6e-02(3.0e-02)	3.0e+00(5.2e-02)	4.1e-01(5.4e-02)	-1.4e-01(3.6e-02)
HDL Cholesterol	MET minutes/week for moderate activity	2.8e-02(1.4e-03)	1.8e-03(1.3e-03)	5.0e-04(1.0e-03)	8.2e-02(1.6e-03)	-2.4e-03(1.4e-03)	2.4e-03(1.2e-03)
HDL Cholesterol	MET minutes/week for vigorous activity	2.8e-02(1.4e-03)	2.0e-04(1.2e-03)	1.0e-03(1.0e-03)	8.1e-02(1.5e-03)	7.0e-04(1.3e-03)	-1.2e-03(1.1e-03)
HDL Cholesterol	Summed MET minutes/week for all activity	2.8e-02(1.4e-03)	6.0e-04(1.2e-03)	5.0e-04(1.0e-03)	8.2e-02(1.5e-03)	-5.0e-04(1.3e-03)	1.6e-03(1.1e-03)

1.Estimation for the multivariate analysis did not converge. Shown are estimates from an univariate analysis, where the full model had a better fit than the null.

2.Estimates are based on standardized data for this trait, due to small phenotypic variance, which resulted in rather small estimates of variance components in absolute terms.

All estimates are derived from analyses of data without a rank-based inverse normal transformation. Standard errors are in brackets. Other model parameters are omitted for simplicity.

Table S7. P-values for comparisons between the full model and nested models for the atherosclerosis risk in communities study.

Main Trait	Lifestlye Covariate	Null vs. Full	GC vs. Full	RC vs. Full
Fibrinogen	Cigarette years of smoking	4.22E-10	1.83E-03	3.53E-01
Fibrinogen	Physical activity:sports domain	3.27E-06	2.57E-03	2.52E-01
Factor VII	Physical activity:sports domain	2.99E-05	7.19E-01	8.62E-01
BMI	Alcohol intake (g/week)	2.60E-12	5.80E-02	1.83E-01
BMI	Physical activity:leisure domain	9.04E-05	3.29E-01	8.69E-01
BMI	Physical activity:sports domain	9.23E-39	3.30E-08	9.60E-01
BMI	Keys score	1.28E-10	1.45E-01	9.73E-01
BMI	Saturated fat intake (g/day)	5.44E-05	4.66E-01	9.75E-01
BMI	Energy from saturated fat (%kcal/day)	6.24E-10	5.85E-02	3.72E-01
BMI	Energy from total fat intake (%kcal/day)	4.09E-08	9.83E-02	3.48E-01
Waist-to-Hip Ratio	Cigarette years of smoking	1.46E-09	1.34E-03	6.44E-01
Waist-to-Hip Ratio	Physical activity:sports domain	5.55E-10	2.35E-05	2.26E-01
Waist-to-Hip Ratio	Total energy intake (kcal/day)	2.00E-05	2.24E-01	4.35E-01
Waist-to-Hip Ratio	Energy from protein (%kcal/day)	7.11E-07	6.29E-04	7.63E-01
Pulse Pressure	Cigarette years of smoking	3.08E-06	2.50E-03	8.35E-01
Diastolic Blood Pressure	Cigarette years of smoking	5.43E-05	3.52E-04	5.93E-02
Heart Rate	Cigarette years of smoking	1.25E-06	1.37E-02	6.59E-01
Heart Rate	Physical activity:leisure domain	1.01E-04	8.47E-03	2.05E-01
Heart Rate	Physical activity:sports domain	3.74E-06	6.20E-06	4.18E-01
HDL2 Cholesterol	Cigarette years of smoking	1.70E-09	4.02E-02	6.94E-01
HDL2 Cholesterol	Physical activity:leisure domain	2.73E-12	1.32E-01	5.20E-01
HDL2 Cholesterol	Carbohydrate intake (g/day)	1.68E-07	3.42E-01	5.77E-01
HDL2 Cholesterol	Total energy intake (kcal/day)	2.40E-08	2.45E-01	4.50E-01
HDL2 Cholesterol	Energy from protein (%kcal/day)	3.69E-08	7.33E-02	6.54E-01
HDL3 Cholesterol	Alcohol intake (g/week)	3.31E-05	2.45E-02	1.88E-01
HDL3 Cholesterol	Physical activity:sports domain	9.68E-06	1.16E-03	1.52E-01
HDL Cholesterol	Physical activity:sports domain	8.85E-06	9.91E-05	4.71E-02
HDL Cholesterol	Monounsaturated fatty acid intake (g/day)	3.17E-05	4.31E-01	9.37E-01
HDL Cholesterol	Energy from protein (%kcal/day)	4.71E-06	7.86E-02	6.00E-01
Apolipoprotein AI	Polyunsaturated fatty acid intake (g/day)	2.94E-05	2.14E-02	2.13E-01
White Blood Cell Count	Cigarette years of smoking	2.00E-65	2.53E-21	1.25E-08
White Blood Cell Count	Physical activity:leisure domain	7.62E-07	5.19E-01	4.36E-01
White Blood Cell Count	Physical activity:sports domain	9.62E-05	1.72E-01	3.82E-01
White Blood Cell Count	Keys score	7.56E-05	8.37E-01	2.32E-01

Only shown for analyses where the full versus null model comparison was significant. Nested models include the null, G-C only and R-C only models. All analyses are based on data after a rank-based inverse normal transformation.

Table S8. P-values for comparisons between the full model and nested models for the UK Biobank.

Main Trait	Lifestlye Covariate	Null vs. Full	GC vs. Full	RC vs. Full
BMI	Alcohol intake (glass & pint/week)	1.19E-14	0.00E+00	0.00E+00
BMI	MET minutes/week for walking	1.49E-16	4.21E-04	6.82E-02
BMI	MET minutes/week for moderate activity	2.05E-29	6.34E-13	8.26E-01
BMI	MET minutes/week for vigorous activity	1.01E-46	8.91E-20	4.74E-01
BMI	Summed MET minutes/week for all activity	5.51E-47	3.16E-15	2.71E-01
BMI	estimated saturated fat intake	6.44E-03	6.37E-01	3.85E-01
Diastolic Blood Pressure	Pack years adult smoking as proportion of life span exposed to smoking	3.03E-04	1.75E-04	1.08E-02
Pulse Pressure	Pack years adult smoking as proportion of life span exposed to smoking	7.13E-06	8.66E-02	4.10E-01
Heart Rate	Pack years adult smoking as proportion of life span exposed to smoking	3.44E-38	5.01E-05	3.92E-01
Heart Rate	MET minutes/week for walking	3.81E-02	6.38E-02	6.96E-01
Heart Rate	Summed MET minutes/week for all activity	3.62E-02	1.05E-01	9.20E-01
Waist-to-Hip Ratio	Pack years adult smoking as proportion of life span exposed to smoking	5.35E-87	6.21E-15	7.46E-01
Waist-to-Hip Ratio	MET minutes/week for moderate activity	9.00E-04	2.07E-01	2.21E-01
Waist-to-Hip Ratio	MET minutes/week for vigorous activity	1.13E-04	2.46E-01	3.12E-01
Waist-to-Hip Ratio	Summed MET minutes/week for all activity	2.23E-04	4.45E-01	2.65E-01
Waist-to-Hip Ratio	estimated total energy intake	2.40E-02	1.81E-02	1.18E-01
White Blood Cell Count	Pack years adult smoking as proportion of life span exposed to smoking	1.25E-138	3.58E-21	6.04E-01
HDL Cholesterol	MET minutes/week for moderate activity	8.01E-08	2.77E-01	3.27E-01
HDL Cholesterol	MET minutes/week for vigorous activity	4.37E-04	5.01E-01	7.39E-01
HDL Cholesterol	Summed MET minutes/week for all activity	5.91E-06	7.65E-01	6.91E-01

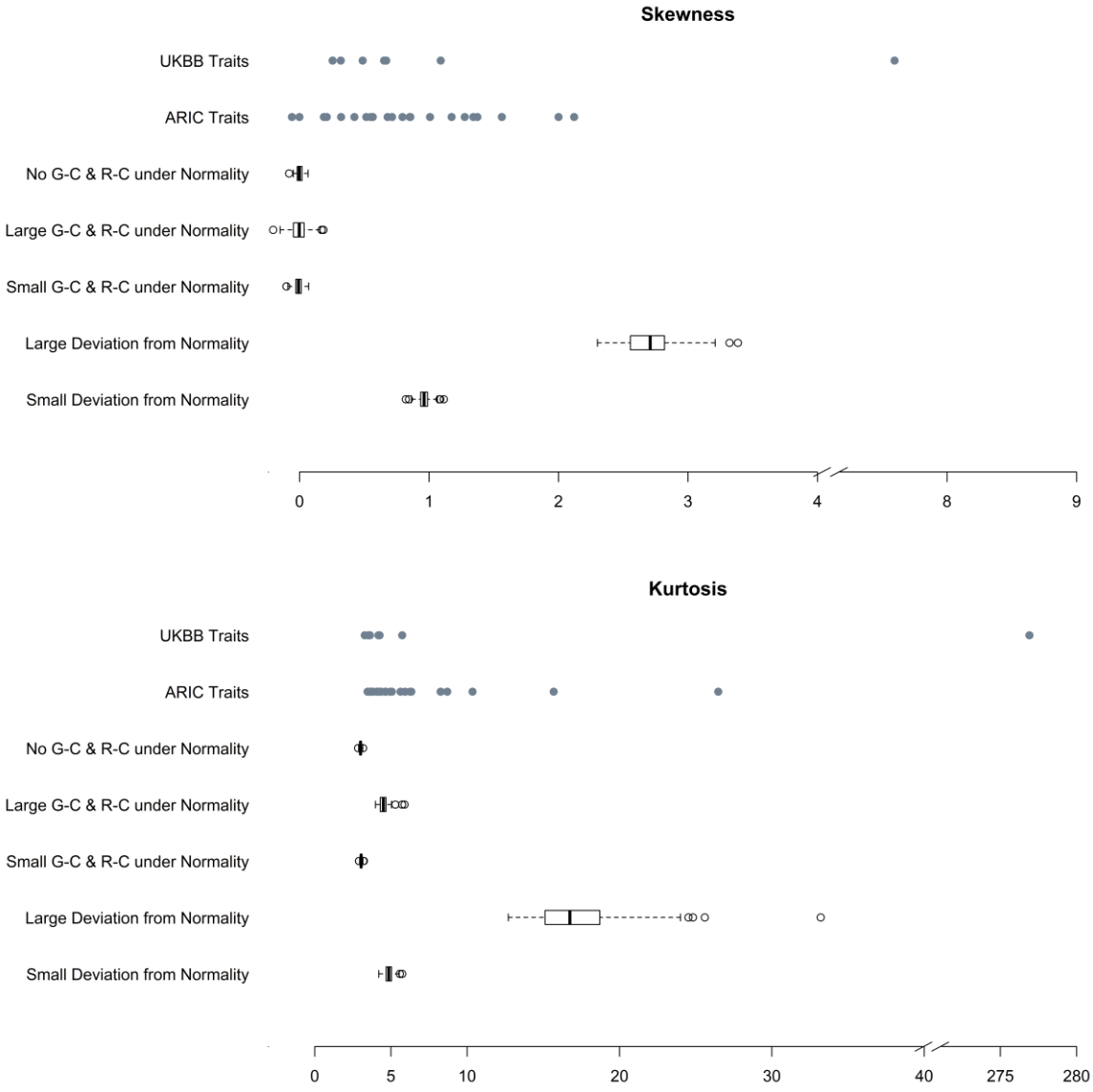
Only shown for analyses where the full versus null model comparison was significant. Nested models include the null, G-C only and R-C only models. All analyses are based on data after a rank-based inverse normal transformation.

Table S9. Genomic relationship within and between top and bottom groups, stratified by G-C interaction estimate, relative to the grand mean genomic relationship.

Main Trait	Lifestyle	Within Top Group		Within Bottom Group		Between Top & Bottom Groups	
		Δ%[1]	*p-value*[2]	Δ%	*p-value*	Δ%	*p-value*
HDL Cholesterol	Physical activity:sports domain	63.4	9.47E-72	77.0	1.46E-105	-69.5	3.44E-165
HDL3 Cholesterol	Physical activity:sports domain	64.3	2.44E-73	67.4	3.47E-81	-66.0	2.72E-148
White Blood Cell Count	Physical activity:sports domain	62.4	2.81E-69	61.4	5.66E-67	-64.0	3.12E-139
HDL2 Cholesterol	Energy from prot1 (%kcal/day)	84.3	2.20E-125	87.8	4.48E-136	-82.6	4.90E-233
White Blood Cell Count	Physical activity:leisure domain	67.3	5.40E-81	66.1	2.03E-77	-65.2	4.46E-145
BMI	Saturated fatty acid intake (g/day)	66.1	4.45E-80	94.0	1.40E-148	-76.3	5.44E-201
HDL3 Cholesterol	Alcohol intake (g/week)	55.1	1.18E-54	57.0	1.71E-58	-56.5	1.72E-109
Heart Rate	Physical activity:sports domain	55.2	1.72E-54	53.2	6.04E-52	-54.1	1.72E-101

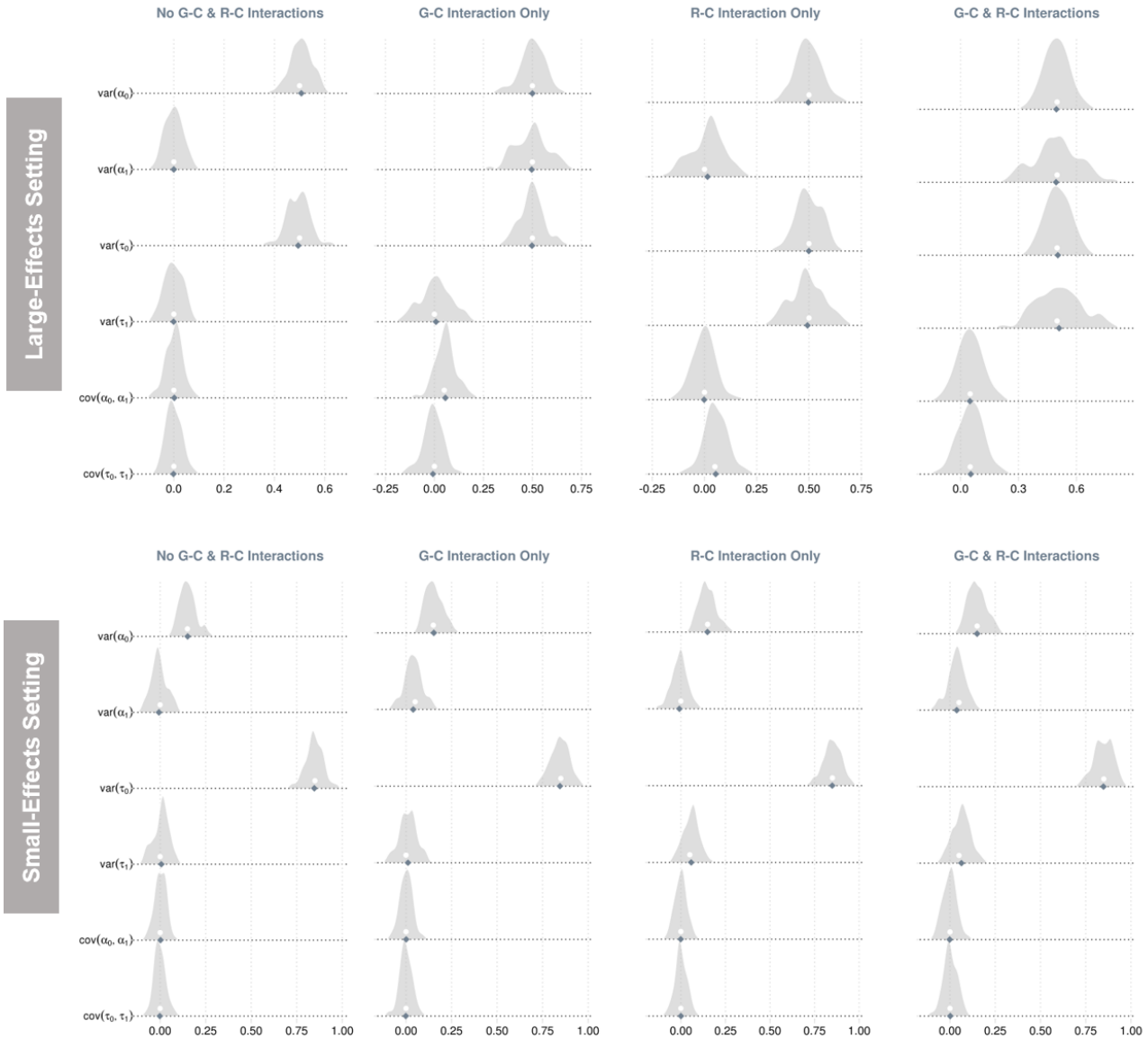
1. Δ% = (group mean - grand mean)/grand mean x 100%; 2. P-values are for two-sided independent t-tests that compare group means with the grand mean

Figure S1. Skewness (top) and kurtosis (bottom) of cardiovascular traits from the atherosclerosis risk in communities study (ARIC) and UK biobank (UKBB).



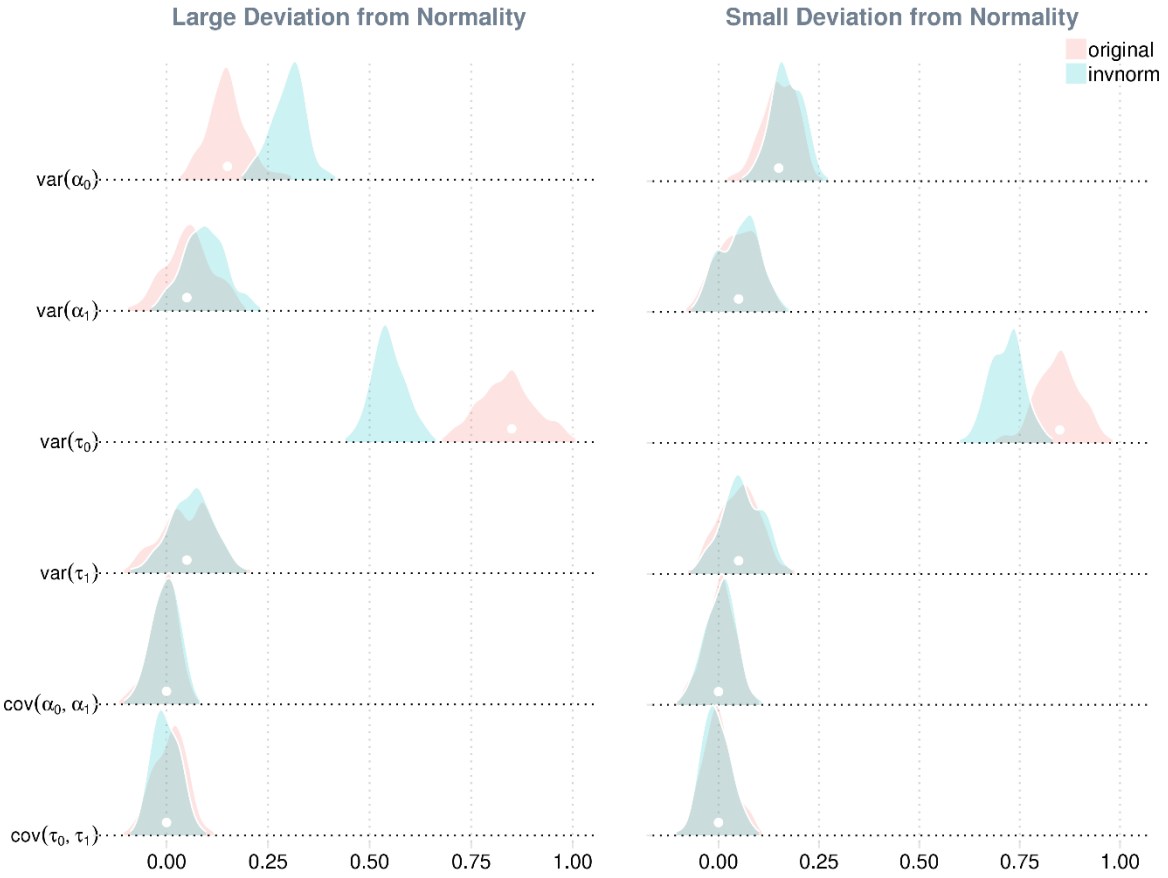
In each panel, categories along the y axis from the top to bottom are traits from the UK biobank, traits from the ARIC study, traits (100 replicates) simulated from the null model, from the full model with large and small effect sizes, and from the full model with non-normal residuals that result in large and small phenotypic deviations from normality. The last five categories are included as references to indicate expected skewness and kurtosis when the model assumption of normality is and is not met.

Figure S2. Sampling distributions of parameter estimates from the full model.



Data for a main trait and a covariate were simulated using 4 multivariate reaction norm models under 2 parameter settings for each model, which together gave arise to 8 combinations of simulation scenarios. The four models were no Genotype-Covariate (G-C) and Residual-Covariate (R-C) interactions (i.e., a null model), G-C interaction only (i.e., a G-C model), R-C interaction only (i.e., a R-C model), and both G-C and R-C interactions (i.e., a full model). The two parameter settings were large and small effect sizes in terms of heritability, G-C and R-C interactions. Each simulation was repeated 100 times, resulting in 100 replicates of simulated data under each scenario. Parameter estimates were obtained from fitting the full model. Shown distributions are for model parameters (i.e., variance & covariance terms) pertaining to the main trait only. True parameter values are shown in dots and means of sampling distributions in diamonds.

Figure S3. Impact of rank-based inverse normal transformation on parameter estimates when the normality assumption is violated.

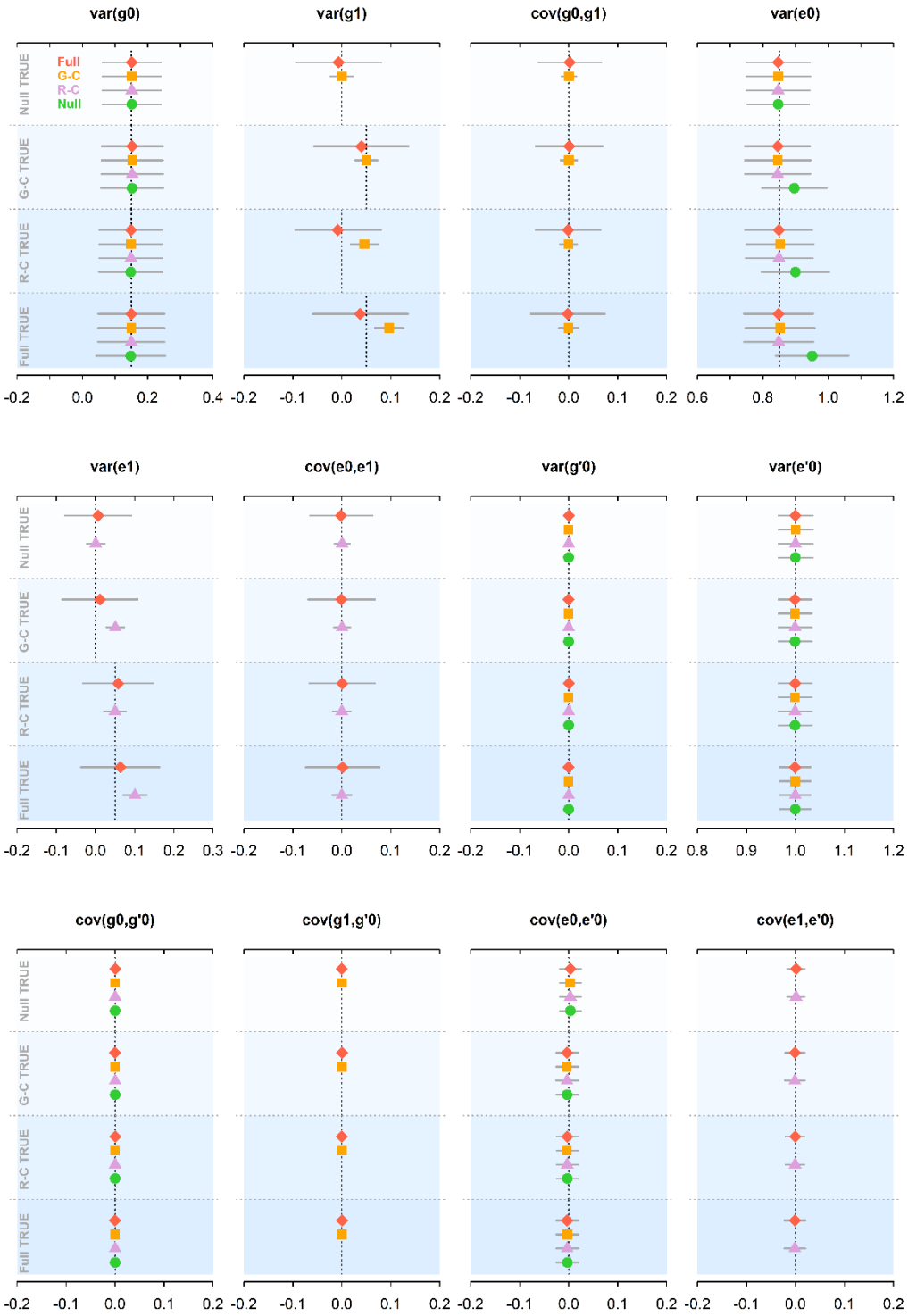


Data for a main trait and a covariate were simulated using a multivariate reaction norm model that included both Genotype-Covariate and Residual-Covariate interactions (i.e., a full model) under two parameter settings, where residuals of the main trait were drawn from distributions that deviated from a normal distribution to different degrees (i.e., small vs. large deviation from normality). Each simulation was repeated 100 times, resulting in 100 replicates of simulated data for each setting. Parameter estimates were obtained from fitting a full model—that assumes normality of all random effects including residuals—to the simulated data before and after a rank-based inverse normal distribution ('original' vs. 'invnorm'). Shown distributions are for model parameters (i.e., variance & covariance terms) pertaining to the main trait only. True parameter values are shown in dots.

Figure S4. Impact of rank-based inverse normal transformation on parameter estimates when normality assumption is met.

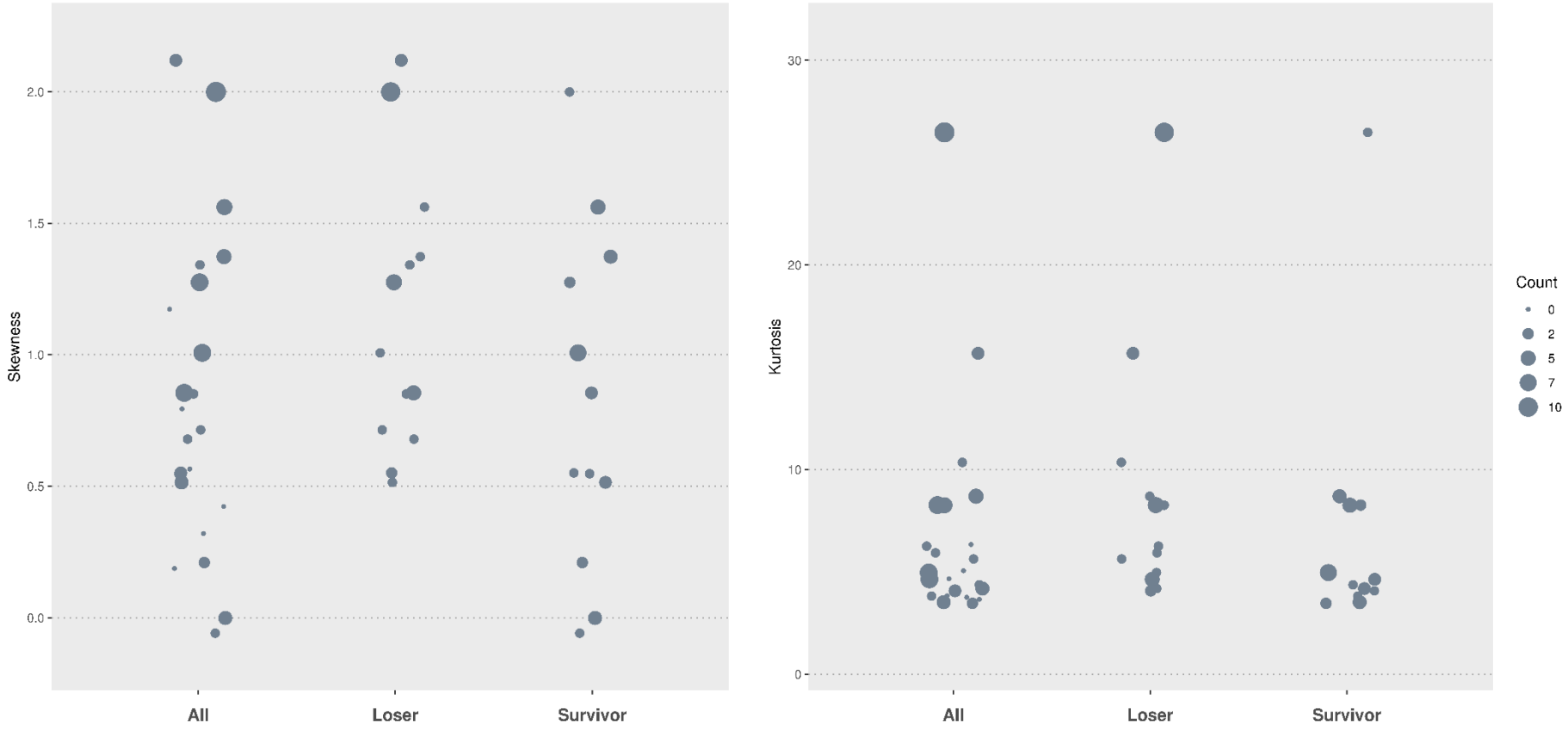
Data for a main trait and a covariate were simulated using a multivariate reaction norm model that included both Genotype-Covariate (G-C) and Residual-Covariate (R-C) interactions (i.e., a full model) under two parameter settings that varied in effect sizes in terms of heritability, G-C and R-C interactions. In each setting, all random effects were drawn from normal distributions. Each simulation was repeated 100 times, resulting in 100 replicates of simulated data for each setting. Parameter estimates were obtained from fitting a full model—that assumes normality of random effects—to the simulated data before and after a rank-based inverse normal distribution ('original' vs. 'invnorm'). Shown distributions are for model parameters (i.e., variance & covariance terms) pertaining to the main trait only. True parameter values are shown in dots.

Figure S5. Sampling distributions of parameter estimates for four multivariate reaction norm models under four simulation scenarios for the small-effects parameter setting.



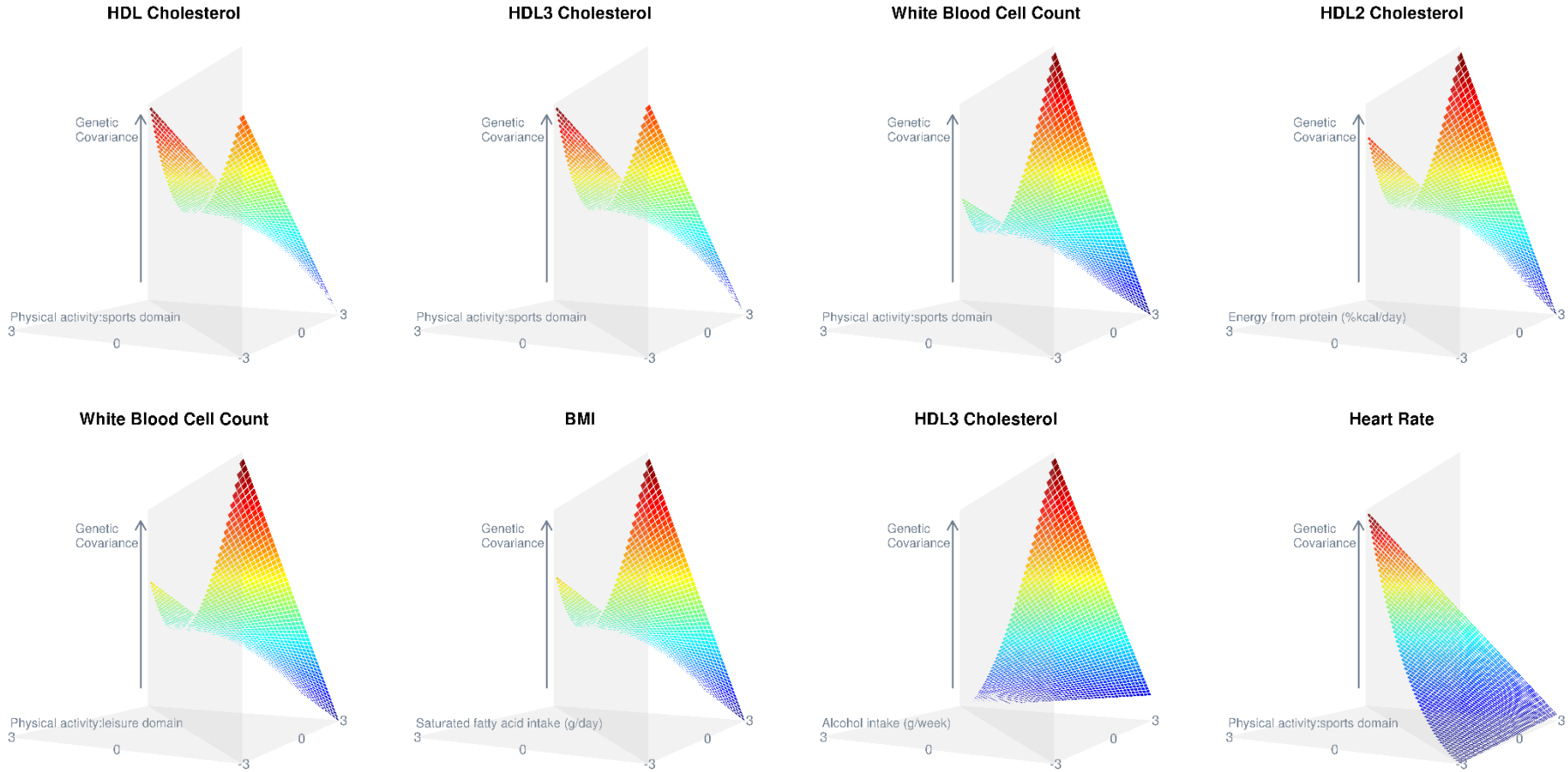
The four scenarios are no Genotype-Covariate (G-C) and Residual-Covariate (R-C) interactions (i.e., a null model), G-C interaction only (i.e., a G-C model), R-C interaction only (i.e., a R-C model), and both G-C and R-C interactions (i.e., a full model). There are 100 replicates under each scenario, with the true value of each parameter being indicated by a vertical dashed line.

Figure S6. Skewness (left panel) and kurtosis (right) of cardiovascular traits from the ARIC dataset by survivorship of rank-based inverse normal transformation.



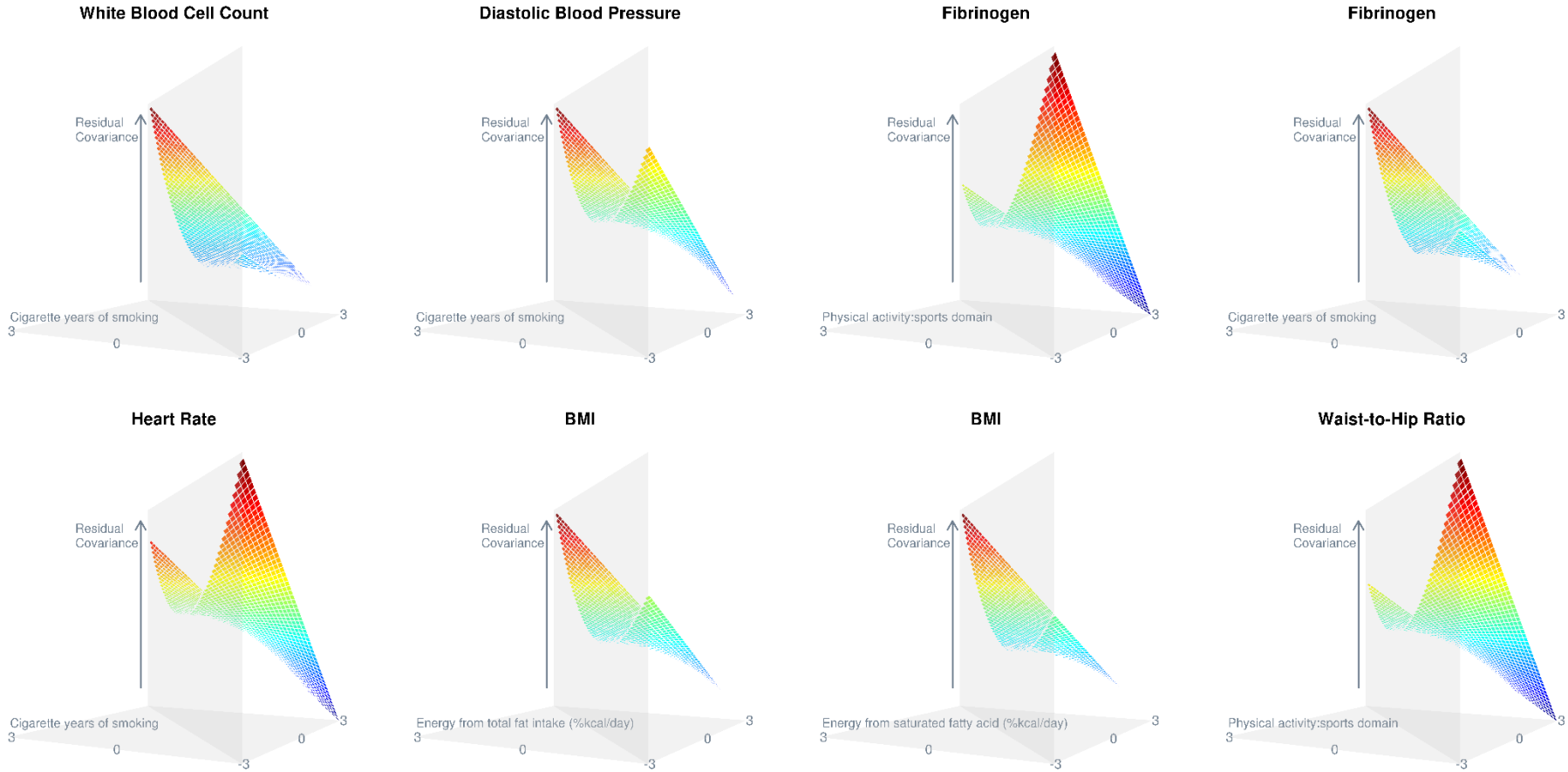
Categories from left to right in each panel are all traits (labelled as “all”), trait that lost one or more signals after a rank-based inverse normal transformation (“loser”), and traits that still had one or more signals after the transformation (“survivor”). Point size is proportional to the count of signals. To reduce overlaps, points are jittered randomly in the horizontal direction. Note that a trait can be both a loser and a survivor.

Figure S7. Estimated genetic covariance with respect to lifestyle covariate.



Both x and y coordinates of each plot cover three standard deviations from the mean of a given lifestyle covariate. The horizontal plane thus represents pairwise combinations of lifestyle covariate values. The corresponding genetic covariance matrix of these combinations, estimated from the full model, is shown as the surface in each plot. The lower triangular part of each matrix, which is identical as the upper triangular part, is removed for simplicity. Diagonal entries of each matrix, shown as the intersection of the surface with the diagonal plane, are estimated genetic variances. Only traits with the first eight largest variance estimates of Genotype-Covariate interaction (see Figure 2 left) are shown. Arrows point to higher values.

Figure S8. Estimated residual covariance with respect to lifestyle covariate.



Both x and y coordinates of each plot cover three standard deviations from the mean of a given lifestyle covariate. The horizontal plane thus represents pairwise combinations of lifestyle covariate values. The corresponding residual covariance matrix of these combinations, estimated from the full model, is shown as the surface in each plot. The lower triangular part of each matrix, which is identical as the upper triangular part, is removed for simplicity. Diagonal entries of each matrix, shown as the intersection of the surface with the diagonal plane, are estimated residual variances. Only traits with the first eight largest variance estimates of Residual-Covariate interaction (see Figure 2 right) are shown. Arrows point to higher values.