Dominant Genetic Variation and Missing Heritability for Human Complex Traits: Insights from Twin versus Genome-wide Common SNP Models

Xu Chen,[1] Ralf Kuja-Halkola,[1] Iffat Rahman,[2] Johannes Arpegård,[3,4] Alexander Viktorin,[1] Robert Karlsson,[1] Sara Hägg,[1] Per Svensson,[3,4] Nancy L. Pedersen,[1] and Patrik K.E. Magnusson[1,*]

In order to further illuminate the potential role of dominant genetic variation in the "missing heritability" debate, we investigated the additive (narrow-sense heritability, h^2) and dominant (δ^2) genetic variance for 18 human complex traits. Within the same study base (10,682 Swedish twins), we calculated and compared the estimates from classic twin-based structural equation model with SNP-based genomic-relatedness-matrix restricted maximum likelihood [GREML(d)] method. Contributions of δ^2 were evident for 14 traits in twin models (average $\delta^2_{twin} = 0.25$, range 0.14–0.49), two of which also displayed significant δ^2 in the GREMLd analyses (triglycerides $\delta^2_{SNP} = 0.28$ and waist circumference $\delta^2_{SNP} = 0.19$). On average, the proportion of h^2_{SNP}/h^2_{twin} was 70% for ADE-fitted traits (for which the best-fitting model included additive and dominant genetic and unique environmental components) and 31% for AE-fitted traits (for which the best-fitting model included additive genetic and unique environmental components). Independent evidence for contribution from shared environment, also in ADE-fitted traits, was obtained from self-reported within-pair contact frequency and age at separation. We conclude that despite the fact that additive genetics appear to constitute the bulk of genetic influences for most complex traits, dominant genetic variation might often be masked by shared environment in twin and family studies and might therefore have a more prominent role than what family-based estimates often suggest. The risk of erroneously attributing all inherited genetic influences (additive and dominant) to the h^2 in too-small twin studies might also lead to exaggerated "missing heritability" (the proportion of h^2 that remains unexplained by SNPs).

Introduction

Heritability is a concept used to denote the relative importance of genetic influences to variability of diseases or complex traits and is loosely defined as the proportion of the phenotypic variance attributed to inherited genetic effects.[1] Several methods can be used to estimate heritability. They are based either on modeling of family correlations in related subjects[2] (distributions of trait similarities among various types of relatives) or on molecular measurements in related or unrelated subjects.[3,4] The classic twin study, often implemented using structural equation modeling (SEM), is the most commonly used family-based approach. Observed intra-pair correlations among genetically identical, monozygotic (MZ) twins and fraternal, dizygotic (DZ) twins are contrasted in order to partition the phenotypic variance into additive (A) genetic variance—so called narrow-sense heritability (h^2), dominant genetic variance (D), and shared (C) and non-shared (E) environmental variance.[2,5] The sum of additive and dominant genetic proportions of variance is often referred to as the broad-sense heritability. As in any family-based modeling, classic twin studies rely on certain important assumptions, the most debated being that MZ and DZ twins share their raising environment to the same extent.

A further complication in the classic twin model is that C and D cannot be estimated simultaneously. This is because the model is under-informed to allow quantification of more than one source to deviance from pure additivity, even if it exists. With this follows that whenever D is indicated in the twin model, it does lend support to *contribution* from D, but the *magnitude* will represent the total net deviance from a pure additive genetic model. Positive contributions to this deviance will stem from dominance (interactions between alleles within the same locus), epistasis (interactions between different loci), as well as other types of higher-order interactions, whereas negative contributions will arise from shared environmental factors. Thus, contributions from both C and D components might very well coexist but "mask" each other, so that the net effect appears as contribution from neither.

Recent methodological developments offer alternatives to estimate heritability via SNPs. When restricting the modeling to include only significantly associated loci identified from genome-wide association studies, it typically accounts for a minute proportion of the h^2 estimated from twin or pedigree studies, a phenomenon originally denoted "missing heritability."[6] By extending the models to utilize contributions from all common SNPs, SNP-based methods like genomic-relatedness-matrix restricted maximum likelihood (GREML) algorithm implemented in genome-wide complex trait analysis (GCTA) can detect considerable shares of the h^2 (typically ~30%–50%).[4,7]

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden; [2]Institute of Environmental Medicine, Karolinska Institutet, Stockholm 17177, Sweden; [3]Department of Medicine-Solna, Karolinska Institutet, Stockholm 17177, Sweden; [4]Department of Emergency Medicine, Karolinska University Hospital, Stockholm 17177, Sweden
*Correspondence: patrik.magnusson@ki.se
http://dx.doi.org/10.1016/j.ajhg.2015.10.004.

The remainder is what now usually is considered to make up the "missing heritability."

Recently, Zhu et al. estimated dominant genetic variance (δ^2) for human complex traits, by applying an extension of GREML algorithm, called GREMLd.[8] The authors observed significant contributions of δ^2 in subsets of traits and samples and estimated the global average contribution to be $1/5^{th}$ of the contribution from A and therefore concluded that dominant genetic variation contributes little to the "missing heritability." Here we investigate the heritability of 18 robustly measured human complex traits including blood biomarkers of cardiovascular disease, kidney function, and diabetes mellitus, as well as three anthropometric reference traits. Our aim was to further illuminate the potential role of dominant genetic variation in the "missing heritability" discussion, by comparing the estimates from both twin-based SEM and SNP-based GREML(d) within the same study base (10,682 Swedish twins).

Material and Methods

Study Population

The subjects of this study (n = 10,682) have all participated in the TwinGene project,[9] a Swedish population-based cohort of twins born between 1911 and 1958. Their average age at phenotypic measurements was 65 (±8) years, and the participants had previously taken part in a computer-assisted telephone interview, Screening Across the Lifespan Twin study (SALT), undertaken between 1998 and 2002. Both of these projects were approved by the local ethics committee at Karolinska Institutet, and all participants have given their informed consent. Zygosity was determined by DNA markers (57% of the study sample) or self-reported childhood resemblance. There were 2,499 monozygotic, 4,154 same-sex dizygotic (SSDZ), and 4,029 opposite-sex dizygotic (OSDZ) twins, totalling 5,074 (48%) men and 5,608 women. Descriptive statistics are shown in Table S1.

Trait Measurements

All physically measured quantitative phenotypes available to all participants in the TwinGene project were investigated. Blood was collected after overnight fasting at a local health-care facility in the morning from Monday to Thursday, to ensure that the tubes with serum would be sent to Karolinska Institutet Biobank before the weekend by overnight post. Samples were stored at −80°C awaiting clinical chemistry and immunological assays. Total cholesterol, triglyceride, low- and high-density lipoproteins, apolipoprotein A1 and B, hemoglobin, C-reactive protein, and glucose were measured by routine methods on semi-automated biochemistry analyzer (Beckman Coulter). Glycosylated hemoglobin A1c was measured by ion exchange chromatography; immunoglobulin A was measured by a reverse-phase protein microarray; cystatin C was measured by particle reinforced immune-turbidimetric analysis using Architect ci8200 immunoassay analyzer; creatinine was measured by an enzymatic method through Arcitect c8000 and Arcitect c16000 (Abbott); and glomerular filtration rate was calculated as $79.901 \times (\text{cystatin C mg/l})^{-1.4389}$. Height, weight, and waist circumference were measured without shoes and in light clothing. Body mass index (BMI) was calculated as $\text{BMI} = \text{weight(kg)/height(m)}^2$. The unit and distribution of each trait in different gender and zygosity subgroups are reported in Tables S2 and S3.

Genotyping

For each individual, 7 ml whole blood was collected in an EDTA tube and genomic DNA was extracted by Puregene extraction kit (Gentra Systems) and subsequently stored at −20°C. Subjects with DNA concentration less than 20 ng/μl, as well as a set of 302 female monozygotic pairs participating in a previous genome-wide genotyping effort, were excluded. DNA from all remaining dizygotic individuals and from one twin within each available monozygotic twin pair (in total, n = 9,896) were sent for genotyping with Illumina OmniExpress bead chip (700K). Quality control was performed and exclusions of samples and SNPs were done according to the following criteria: genotype missingness > 0.03, individual missingness > 0.03, minor allele frequency < 0.01, Hardy-Weinberg equilibrium p value $< 10^{-7}$, sex mismatch, heterozygosity (individuals with an F-statistic larger than five standard deviations from the sample mean), cryptic (unknown) relatedness, or phenotypic information missing on more than five traits. Finally, 9,606 individuals and 644,556 SNPs remained.

Data Handling

Data handling, descriptive statistics, covariate adjustment, and normalization were performed in SAS v.9.4 (SAS Institute). The difference in means between males and females was tested for each trait by t test. Raw values of each trait were adjusted for age, sex, and the first ten principal components based on genotypes (9,617 individuals and 644,556 SNPs that passed genotyping QC) in linear regression models, then residuals from the regression were rank order normalized, resulting in standard normal distributions.

Twin-Based SEM

Twin order was randomly assigned, singletons and pairs with missing values for more than five traits were removed, and finally 3,870 complete twin pairs (1,088 MZ, 1,443 SSDZ, and 1,339 OSDZ) were used in twin-based analyses. Structural equation modeling of the observed covariance in MZ and DZ twin pairs was performed to find maximum likelihood estimates for additive genetic effects (A; the sum of the effects of individual loci), dominant genetic effects (D; interactions between alleles within the same locus), common/shared environmental effects (C; contributes to the similarities between relatives living together), and unique/non-shared environmental effects (E; specific to individual, contributes to the dissimilarities between family members), contributing to the variance within, and covariance between, individuals for each phenotype. Akaike information criterion[10] was used to compare the goodness of fit of ACE (a model including A, C, and E), ADE (including A, D, and E), and AE (including A and E) models to find the most parsimonious model. The narrow- and broad-sense heritability (h^2 and H^2) was estimated, corresponding to the proportion of phenotypic variance attributable to additive genetic variance, or additive plus dominance genetic variances, respectively. All twin-based analyses were performed with the OpenMx package[11] in R.

SNP-Based GREML(d)

GREML(d) was implemented in GCTA to estimate h^2 and δ^2 via comparison of empirical genetic resemblance of unrelated

individuals, based on identity-by-state when all genome-wide common SNPs are fitted as random effects in a mixed linear model. One twin in each complete dizygotic pair (both same-sex and opposite-sex) was randomly removed from the 9,606 individuals with both genotypes and phenotypes available. The remaining 6,812 individuals were used in the first step of generating pairwise additive genomic relationship matrix. Subsequently, one individual within each more distantly related pair (cut-off value > 0.025, corresponding to a relatedness between second and third cousins) was removed and dominant genomic relationship matrix was generated based on these 5,779 unrelated individuals (1,185 MZ, 2,316 SSDZ, and 2,278 OSDZ twins). For each phenotype, restricted maximum likelihood was used to estimate the variance explained by all SNPs. Both the SNP-based additive genetic variance, the so-called "chip heritability" (h^2_{SNP}), and dominant genetic variance (δ^2_{SNP}) was estimated.

Contact Frequency and Age at Separation within Twin Pairs

Self-reported intra-pair contact frequency (the frequency by which the twins in a pair said they met in person) and age at separation from the co-twin constitutes independent measures of degree of shared environment within twin pairs. Such measures were available from the SALT interview for more than 90% of the complete twin pairs. The replies to questions about how often the participants usually met with their co-twin were divided into four levels: (1) less than once a year, (2) on a yearly basis, (3) on a monthly basis, or (4) on a weekly basis; the measure is here used as a continuous variable. When answers were available from both twins, the within-pair average value was calculated and used for each pair. Violation of the equal environment assumption was tested by comparing means (t test) between MZ and DZ twins in contact frequency/age at separation. To test whether the degree of shared environment was related to within-pair trait similarity, the correlation between contact frequency/age at separation and absolute intra-pair difference in adjusted trait levels was examined among MZ twins. MZ intra-pair correlation stratified by level of shared environment was also estimated for each trait. For contact frequency, we divided all pairs into low (defined as ≤ 3, on the monthly basis) and high (defined as >3, more than monthly) groups, and for age at separation we divided the pairs by the median value.

Results

The classic twin model indicated contributions of dominant genetic effects (ADE was the best-fitting model) for 14 out of the 18 traits, with an average δ^2_{twin} of 0.25 (Table 1). In two of these traits (triglycerides and waist circumference), we also observed significant dominance in the SNP-based GREMLd model. Notably, the significant δ^2_{SNP} of waist circumference observed in the ARIC cohort from the paper of Zhu et al.[8] was successfully replicated in our data ($\delta^2_{SNP} = 0.19$, 95% CI 0.01–0.37, $p = 0.01$). The large estimate of δ^2_{SNP} observed for triglycerides was not seen in the corresponding δ^2_{twin}, possibly due to chance in the sampling and the fact that the GREML(d) and the twin-based models are independent methods since they rely on different contrasts. GREMLd also indicated contributions from dominance for six additional traits, but all were non-significant (Table S4), possibly because the sample size of unrelated individuals was too small for sufficient power.

In the classic twin analyses, the AE model was the best-fitting model for high-density lipoprotein and apolipoprotein A1, whereas the ACE model was the best-fitting model for immunoglobulin A and height (Table S5). For the h^2_{SNP} estimated from GREML, the means were very similar between the ADE- and AE-fitted traits, equal to 0.22 and 0.21, respectively. This was not supported by results from the twin model, in which the average h^2_{twin} of ADE-fitted traits was half as big as the average h^2_{twin} of AE-fitted traits.

The proportion of h^2_{SNP}/h^2_{twin} indicates how much of the A component estimated from the twin-based model is explained by the common SNP-based model. A large proportion, on average 70% of h^2_{twin}, was captured by h^2_{SNP} for ADE-fitted traits, whereas for the two AE-fitted traits, h^2_{SNP} explained only 31% of the h^2_{twin}, similar to the proportions generally reported in previous studies. If previous studies were underpowered to identify significant dominance components, they might instead have attributed it to the additive component in a more parsimonious AE model. This is similar to the finding for high-density lipoprotein in the current study; the pattern of the intra-pair correlations ($r_{MZ} > 2r_{DZ}$) indicated presence of dominance (Table S6), but we appear underpowered to declare it significant. If we mimic this situation and adopt AE models for all the D-influenced traits (i.e., let A be the sum of A and D components), the average value of h^2_{SNP}/H^2_{twin} would decrease to 39%. For the sake of completeness, we also performed a sex-limitation SEM that estimated the heritability by gender, but by doing so the power to identify dominant effects decreased and AE model became the best-fitted model for several traits (Tables S7 and S8). For creatinine and GFR, there are pronounced differences in variance components estimates between males and females, which is in agreement with a previous report from the same study base.[12]

Using self-reported measures of degree of shared environment with MZ co-twin, we found independent evidence for influences from shared environment for a subset of traits. Self-reported contact frequency and the number of years spent together before separation were both significantly higher for MZ than for SSDZ and OSDZ twins (Table S9). Together these results indicate violations of the equal environment assumption as a potential problem. In order to get a sense of the magnitude of bias such a violation might be associated with, we calculated the MZ intra-pair correlations stratified by level of shared environment in a high versus a low group (Table S10). Even though the levels of shared environments were considerably larger in the high group, 1.8 standard deviation (SD) for contact frequency and 1.4 SD for age at separation, the trait level similarity was influenced only modestly with r_{MZ} estimates on average ~0.05 larger in the high group.

Contact frequency was weakly albeit significantly related to absolute within MZ-pair difference in high-density lipoproteins, body mass index, weight, and waist circumference

Table 1. Estimates from Twin-Based Structural Equation Model and SNP-Based GREML(d)

	Twin-Based SEM[a]							SNP-Based GREML(d)[b]					
Trait	r_{MZ}	r_{DZ}	Best M	h^2_{twin}	95% CI	δ^2_{twin}	95% CI	h^2_{SNP}	95% CI	δ^2_{SNP}	95% CI	h^2_{SNP}/h^2_{twin}	h^2_{SNP}/H^2_{twin}[c]
TC	0.48	0.19	ADE	0.28	(0.13,0.43)	0.19	(0.03,0.36)	0.15	(0.03,0.27)	0.00	(0.00,0.18)	54%	32%
LDL	0.46	0.18	ADE	0.23	(0.08,0.38)	0.24	(0.07,0.41)	0.16	(0.04,0.28)	0.00	(0.00,0.18)	70%	34%
Apolipoprotein B	0.52	0.23	ADE	0.39	(0.25,0.53)	0.14	(0.00,0.30)	0.14	(0.02,0.26)	0.00	(0.00,0.18)	36%	26%
Triglyceride	0.55	0.24	ADE	0.42	(0.27,0.55)	0.14	(0.00,0.30)	0.31	(0.19,0.43)	0.28	(0.10,0.46)[d]	74%	55%
C-reactive protein	0.42	0.19	ADE	0.30	(0.15,0.44)	0.14	(0.00,0.31)	0.37	(0.25,0.49)	0.00	(0.00,0.18)	123%	84%
Glucose	0.51	0.20	ADE	0.24	(0.09,0.38)	0.30	(0.15,0.46)	0.17	(0.05,0.29)	0.15	(0.00,0.33)	71%	31%
HbA1c	0.69	0.28	ADE	0.37	(0.24,0.51)	0.35	(0.21,0.49)	0.20	(0.08,0.32)	0.00	(0.00,0.18)	54%	28%
Hemoglobin	0.55	0.24	ADE	0.41	(0.26,0.55)	0.15	(0.00,0.30)[e]	0.21	(0.09,0.33)	0.00	(0.00,0.18)	51%	38%
Cystatin C	0.57	0.26	ADE	0.42	(0.28,0.56)	0.18	(0.03,0.34)	0.27	(0.15,0.39)	0.05	(0.00,0.23)	64%	45%
Creatinine	0.58	0.24	ADE	0.35	(0.21,0.50)	0.24	(0.09,0.40)	0.18	(0.06,0.30)	0.01	(0.00,0.19)	51%	31%
eGFR	0.57	0.24	ADE	0.38	(0.23,0.52)	0.21	(0.05,0.36)	0.32	(0.20,0.44)	0.03	(0.00,0.21)	84%	54%
Body mass index	0.68	0.24	ADE	0.28	(0.13,0.42)	0.41	(0.26,0.56)	0.21	(0.09,0.33)	0.02	(0.00,0.20)	75%	30%
Weight	0.73	0.27	ADE	0.37	(0.23,0.51)	0.35	(0.21,0.50)	0.26	(0.14,0.38)	0.11	(0.00,0.29)	70%	36%
WC	0.63	0.20	ADE	0.15	(0.01,0.29)	0.49	(0.34,0.65)	0.16	(0.04,0.28)	0.19	(0.01,0.37)[d]	107%	25%
ADE-Average	–	–	–	0.33	–	0.25	–	0.22	–	0.06(0.24)[d]	–	70%	39%
HDL	0.67	0.31	AE	0.66	(0.63,0.69)	–	–	0.24	(0.12,0.36)	0.01	(0.00,0.19)	36%	–
Apolipoprotein A1	0.65	0.34	AE	0.66	(0.63,0.68)	–	–	0.17	(0.05,0.29)	0.09	(0.00,0.27)	26%	–
AE-Average				0.66	–	–	–	0.21	–	0.05	–	31%	–
Immunoglobulin A	0.43	0.28	ACE	0.40	(0.29,0.51)	0.07[f]	(0.00,0.15)	0.24	(0.12,0.36)	0.00	(0.00,0.18)	60%	–
Height	0.87	0.48	ACE	0.77	(0.71,0.83)	0.09[f]	(0.04,0.15)	0.62	(0.50,0.74)	0.00	(0.00,0.18)	81%	–
ACE-Average	–	–	–	0.59	–	0.08[f]	–	0.43	–	0.00	–	–	–

Abbreviations are as follows: TC, total cholesterol; WC, waist circumference; LDL and HDL, low- and high-density lipoproteins; HbA1c, glycosylated hemoglobin A1c; eGFR, estimated glomerular filtration rate (machine-based calculation from cystatin C); r_{MZ} and r_{DZ}, coefficients of intra-pair correlations within monozygotic and dizygotic twin pairs; Best M, the best-fitting model for each trait according to Akaike information criterion; h^2_{twin} and δ^2_{twin}, additive and dominant genetic variance estimated from twin model; 95% CI, 95% confidence interval; h^2_{SNP} and δ^2_{SNP}, additive and dominant genetic variation estimated from SNP model.
[a]Estimates from classical twin-based structural equation model (SEM) including 3,870 twin pairs.
[b]Estimates from directly genotyped SNPs of 5,779 unrelated individuals in genomic-relatedness-matrix restricted maximum likelihood [GREML(d)] method.
[c]H^2 indicates broad-sense heritability including both additive (h^2) and dominant (δ^2) genetic variance.
[d]Value in parentheses equals the average of the two significant estimates.
[e]Non-significant.
[f]Shared environmental components estimated in ACE model.

(Table 2). Similar results were obtained when using age at separation as the indicator of degree of shared environment, with significant correlations observed with absolute within MZ-pair difference in high-density lipoproteins, body mass index, and weight. These evaluations are restricted to MZ twins since the aim is to test the relation between degree of shared environment and intra-pair trait similarity *un-confounded* by genetic influences. In DZ pairs the genetic sharing will differ between pairs, which means that the correlations are not straightforward to interpret.

Discussion

Our results from both the classic twin-based and the common SNP-based models lend support to a more prominent role of dominant genetic variation than previous studies generally have reported for similar traits. The large size of the study and advanced age of participants might be contributing factors of importance for this finding. The results also highlight the potential risk of systematic ignorance of deviances from pure additivity in smaller twin studies.

Since heritability by definition is population specific, comparisons of estimates obtained from twin-based and genome-wide common SNP-based models can be achieved reliably only when both types of analyses are performed within the same study base. With the large number of genotyped twin pairs available in TwinGene, it is well suited for such comparisons. We attribute the unusual detection of significant δ^2_{twin} for a majority of traits in our sample to the increased power to discriminate A from D that

Table 2. Correlations between Absolute Intra-pair Difference of Trait Values and Co-twin Contact Frequency and Age at Separation from Co-twin in Monozygotic Twins

	Contact Frequency			Separation Age		
Trait	**r**	**n_{pair}**	**p**	**r**	**n_{pair}**	**p**
Total cholesterol	0.004	1,066	0.90	0.014	1,027	0.66
High-density lipoprotein	*−0.081*	*1,066*	*0.01*	*−0.064*	*1,027*	*0.04*
Low-density lipoprotein	0.017	1,044	0.59	−0.010	1,007	0.74
Apolipoprotein A1	−0.060	1,065	0.05	−0.052	1,026	0.09
Apolipoprotein B	−0.010	1,065	0.75	0.031	1,026	0.32
Triglyceride	−0.047	1,066	0.13	−0.026	1,027	0.40
C-reactive protein	−0.028	1,065	0.35	0.016	1,026	0.62
Glucose	−0.032	1,066	0.29	−0.008	1,027	0.79
Glycosylated hemoglobin A1c	−0.052	1,065	0.09	−0.036	1,026	0.25
Hemoglobin	0.005	1,063	0.87	−0.026	1,024	0.40
Cystatin C	−0.044	1,029	0.16	0.029	993	0.37
Creatinine	−0.035	1,029	0.26	0.019	993	0.55
Glomerular filtration rate	−0.031	1,029	0.32	0.009	993	0.79
Immunoglobulin A	0.008	1,058	0.79	0.002	1,020	0.95
Body mass index	*−0.071*	*1,061*	*0.02*	*−0.089*	*1,022*	*<0.01*
Weight	*−0.085*	*1,064*	*0.01*	*−0.063*	*1,025*	*0.04*
Waist circumference	*−0.083*	*1,063*	*0.01*	−0.048	1,024	0.12
Height	−0.020	1,063	0.52	−0.033	1,024	0.30

Abbreviations are as follows: r, spearman correlation coefficient for the correlation between absolute intra-pair difference and contact frequency/separation age for each trait; p, p values. Significant estimates are in italics.

comes with the large sample of twins of older age. This view is supported by findings from other unusually large population-based studies, such as a recent Dutch twin-family study on blood biomarker levels and metabolic syndrome traits, which detected significant D effects that increased with age.[13]

Because twins represent only a small fraction of the population, the sample size in twin studies is usually limited and most previous studies have included fewer than 1,000 twin pairs, which might provide inadequate power to significantly declare contributions from variance components indicating deviance from additivity (i.e., C and D). Instead, in smaller studies the dominant genetic effects or shared environmental effects are typically attributed to additive genetics in more parsimonious AE models. In a recent very large meta-analysis of estimates from twin studies,[14] the vast majority of investigated traits were reported to be consistent with a simple AE model in which all twin similarity was attributed to A, with the remaining variance explained by non-shared environmental factors. However, close to 50% of all reported joint r_{MZ} and r_{DZ} estimates actually showed a pattern of deviance supporting D ($r_{MZ} > 2r_{DZ}$). Still, D was handled as a part of A (A = D + A) in all such traits. This is similar to what we would observe if we would split the large TwinGene material into several smaller samples; we would then be unable to declare D significant (due to lack of power) and instead report AE as the best-fitting model, and consequently attribute D to the A component.

Twin studies represent the classic design to disentangle the genetic and environmental contributions to familial aggregation/correlation. The relative importance of genetic and environmental effects is estimated by decomposing the total variance into different components (A, C or D, and E). The decomposition relies on important assumptions: MZ share 100% while DZ share on average 50% of their inherited genome, twins within MZ and DZ pairs share the raising environment to the same extent (equal environment assumption), and there is no correlation or interaction between genes and environment.[2] For traits in which the familial aggregation is solely due to additive genetic effects, the MZ correlation is expected to be exactly twice the DZ correlation ($r_{MZ} = 2r_{DZ}$). If there are additional influences from shared environmental effects (C), the additive pattern becomes distorted by making DZ more similar to MZ twins (i.e., $r_{MZ} < 2r_{DZ}$). When the deviance goes in the other direction ($r_{MZ} > 2r_{DZ}$), dominant genetic effects (D) are usually assumed to cause the distortion from pure additivity.

The inability to estimate C and D simultaneously inherent in the classic twin model means that whenever there is a deviance from a pure additive genetic model, the model will provide support to the *existence* of either shared environment or dominance deviance; however, significant contribution of one says nothing about absence or presence of the other. This means that contributions from both components might coexist but "mask" each other, so that the net effect appears as contribution from neither. Thus, when C exists simultaneously with D, it will tend to counterbalance and even outweigh deviance contribution from D. However, since the twins in our sample were relatively old (mean age was 65 years), contributions from shared environmental factors might have been attenuated, leaving D more prone to be observed.

Our data provide independent evidence for simultaneous contribution of C for some traits in which the SEM declare ADE to be the best-fitting model. We conclude this from the fact that there were significant negative correlation between self-reported contact frequency in MZ pairs and absolute within-pair trait difference for high-density lipoprotein and three weight-related traits (Table 2). Further, the observed relation was negative for 14 out of 18 traits, indicating that a general trend might be present also for other traits. Number of years spent together before separation showed similar relations to absolute within-pair trait difference in MZ twins. Twins staying together longer tend to display more similar trait values. The group-mean differences between MZ and DZ twins in degree of shared environment were 0.58 and 0.39 SD for contact frequency and age at separation, respectively. Thus, the core twin model assumption of equal shared environment between MZ

and DZ appears to be violated. One potential consequence of such a violation is that the D component might become inflated in the twin model, while GREML(d) would stay unaffected. This difference could be argued one reason for the markedly larger D component as estimated in the twin-based SEM compared to the GREML(d). However, the relation between degree of shared environment and within-pair trait difference was weak: the strongest correlation found for contact frequency was for weight (−0.085), and for separation age it was BMI (−0.089) (Table 2). The weak relation was apparent also from comparisons of trait correlations in strata of the MZ twins sharing most and least environment (Table S10). The mean difference in level of sharing between the two groups is at least three times larger than the difference observed between MZ and DZ twins. Still, the correlations in trait levels were only very modestly different between the high and low group. Thus, we consider the bias potentially introduced by the violation of the equal environment assumption to be small, and thus not a prominent reason for the discrepancy between twin-based versus GREMLd-based estimation of D.

Another way to obtain independent evidence for contributions of C is to study non-biological relations such as adoptive or step relations. In a previous investigation on military conscription data of BMI at age of 18, significant correlations were observed also among non-biological (step- and adoptive) relatives.[15] This indication of C was supported by significantly stronger correlation in maternal compared to paternal half-brothers, arguably reflecting that children most often follow mothers upon divorce in the studied population, or that mothers have a generally stronger impact on the relevant family environment (eating habits, food choices, etc.) as compared to fathers.[16]

It is clear that the MZ correlation coefficient provides per se an unbiased upper bound of the proportion of variance that genetics (both additive and non-additive) ultimately could explain, but in the contemporary literature there exist different opinions about what should be considered the relevant denominator in the concept of "missing heritability." If it is the broad-sense heritability, an additive modeling of genotypes should not be expected to explain anything but the additive fraction (i.e., we would have to accept that a portion will remain inaccessible). On the other hand, we and others consider the relevant denominator to be the narrow-sense heritability, and the missing heritability to be the proportion of h^2 that remains unexplained by SNPs, equal to $1 - (h^2_{SNP}/h^2_{twin})$.

During the past decade, many explanations of the missing heritability phenomenon have been suggested. Some have focused on using larger numbers of common or rare variants to capture more of the functional genetic variance;[17,18] others have suggested that missing heritability is due to an overestimation of the additive genetic effects because cryptic contribution of epistatic interactions between loci (often termed "I"), creating something denoted "phantom heritability."[19,20] Our results suggest a similar concept, the possibility that h^2 tends to be overestimated if there is inadequate power to discriminate A and D components in the twin model. Letting the A component from a more parsimonious AE model represent the h^2 will provide a value that will be closer to the broad-sense heritability (variation due to A plus variation due to D). There is also a possibility that epistatic interactions are captured differentially by the classic twin-based and the SNP-based GREML(d) heritability estimation and thereby might be responsible for some of the differences between the two models. However, a recent paper suggested that epistatic effects will contribute little to genetic variance in outbred populations.[21]

Even though additive genetics most probably constitutes the bulk of genetic influences to most complex traits, our results from both twin-based and SNP-based models lend support to a more prominent role of dominant genetic variation than what most earlier studies have indicated. We believe simultaneous contributions from both C and D might be a common situation for many traits. Extended twin-family study designs including more family members (e.g., parents, offspring, and non-twin siblings) might offer improved possibilities to verify the existence of D effects,[22,23] but such materials of adequate size are unfortunately exceptionally rare. Previous elegant simulation studies performed in extended twin-family structures[24] lend support for the view we here present that simultaneous presence of C and D might be a common phenomenon despite the fact that classic twin studies rarely find evidence for either. We foresee a future development where integration of twin, family, and molecular-based methods allow more accurate quantification of additive and non-additive proportions of genetic influences, which in turn might help us to reclaim the remains of the missing heritability.

Supplemental Data

Supplemental Data include ten tables with details about samples, methods, descriptive statistics, and results and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2015.10.004.

Acknowledgments

Karolinska Institutet financially supports the Swedish Twin Registry. The study was supported by Swedish Research Council, Swedish Heart-Lung Foundation, and China Scholarship Council.

Received: July 7, 2015
Accepted: October 10, 2015
Published: November 5, 2015

Web Resources

The URLs for data presented herein are as follows:

GCTA-GREML(d), http://cnsgenomics.com/software/gcta/download.html
OpenMx - Advanced Structural Equation Modeling, http://openmx.psyc.virginia.edu/
R statistical software, http://www.r-project.org/

References

1. Tenesa, A., and Haley, C.S. (2013). The heritability of human disease: estimation, uses and abuses. Nat. Rev. Genet. *14*, 139–149.
2. Neale, M.C., and Maes, H.H.M. (2004). Methodology for Genetic Studies of Twins and Families (Dordrecht, Netherlands: Kluwer Academic Publishers B.V), pp. 2–7.
3. Visscher, P.M., Medland, S.E., Ferreira, M.A., Morley, K.I., Zhu, G., Cornes, B.K., Montgomery, G.W., and Martin, N.G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet. *2*, e41.
4. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
5. van Dongen, J., Slagboom, P.E., Draisma, H.H., Martin, N.G., and Boomsma, D.I. (2012). The continuing value of twin studies in the omics era. Nat. Rev. Genet. *13*, 640–653.
6. Maher, B. (2008). Personal genomes: The case of the missing heritability. Nature *456*, 18–21.
7. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.
8. Zhu, Z., Bakshi, A., Vinkhuyzen, A.A., Hemani, G., Lee, S.H., Nolte, I.M., van Vliet-Ostaptchouk, J.V., Snieder, H., Esko, T., Milani, L., et al.; LifeLines Cohort Study (2015). Dominance genetic variation contributes little to the missing heritability for human complex traits. Am. J. Hum. Genet. *96*, 377–385.
9. Magnusson, P.K., Almqvist, C., Rahman, I., Ganna, A., Viktorin, A., Walum, H., Halldner, L., Lundström, S., Ullén, F., Långström, N., et al. (2013). The Swedish Twin Registry: establishment of a biobank and other recent developments. Twin Res. Hum. Genet. *16*, 317–329.
10. DeLeeuw, J. (1992). Breakthroughs in Statistics: Introduction to Akaike (1973) Information Theory and an Extension of the Maximum Likelihood Principle (New York: Springer), pp. 599–609.
11. Neale, M.C., Hunter, M.D., Pritikin, J.N., Zahery, M., Brick, T.R., Kirkpatrick, R.M., Estabrook, R., Bates, T.C., Maes, H.H., and Boker, S.M. (2015). OpenMx 2.0: extended structural equation and statistical modeling. Psychometrika. Published online January 27, 2015. http://dx.doi.org/10.1007/s11336-014-9435-8.
12. Arpegård, J., Viktorin, A., Chang, Z., de Faire, U., Magnusson, P.K., and Svensson, P. (2015). Comparison of heritability of cystatin C- and creatinine-based estimates of kidney function and their relation to heritability of cardiovascular disease. J. Am. Heart Assoc. *4*, e001467.
13. van Dongen, J., Willemsen, G., Chen, W.M., de Geus, E.J., and Boomsma, D.I. (2013). Heritability of metabolic syndrome traits in a large population-based sample. J. Lipid Res. *54*, 2914–2923.
14. Polderman, T.J., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M., and Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat. Genet. *47*, 702–709.
15. Clément, K., and Sørensen, T.I.A. (2007). Obesity: Genomics and Postgenomics (New York: Informa Healthcare), pp. 19–37.
16. Magnusson, P.K., and Rasmussen, F. (2002). Familial resemblance of body mass index and familial risk of high and low body mass index. A study of young men in Sweden. Int. J. Obes. Relat. Metab. Disord. *26*, 1225–1231.
17. Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., and Lander, E.S. (2014). Searching for missing heritability: designing rare variant association studies. Proc. Natl. Acad. Sci. USA *111*, E455–E464.
18. Golan, D., Lander, E.S., and Rosset, S. (2014). Measuring missing heritability: inferring the contribution of common variants. Proc. Natl. Acad. Sci. USA *111*, E5272–E5281.
19. Zuk, O., Hechter, E., Sunyaev, S.R., and Lander, E.S. (2012). The mystery of missing heritability: genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. USA *109*, 1193–1198.
20. Ritchie, M.D. (2015). Finding the epistasis needles in the genome-wide haystack. Methods Mol. Biol. *1253*, 19–33.
21. Mäki-Tanila, A., and Hill, W.G. (2014). Influence of gene interaction on complex trait variation with multilocus models. Genetics *198*, 355–367.
22. Rettew, D.C., Rebollo-Mesa, I., Hudziak, J.J., Willemsen, G., and Boomsma, D.I. (2008). Non-additive and additive genetic effects on extraversion in 3314 Dutch adolescent twins and their parents. Behav. Genet. *38*, 223–233.
23. Keller, M.C., Coventry, W.L., Heath, A.C., and Martin, N.G. (2005). Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. Behav. Genet. *35*, 707–721.
24. Keller, M.C., Medland, S.E., and Duncan, L.E. (2010). Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. Behav. Genet. *40*, 377–393.

The American Journal of Human Genetics

Supplemental Data

Dominant Genetic Variation and Missing Heritability for Human Complex Traits: Insights from Twin versus Genome-wide Common SNP Models

Xu Chen, Ralf Kuja-Halkola, Iffat Rahman, Johannes Arpegård, Alexander Viktorin, Robert Karlsson, Sara Hägg, Per Svensson, Nancy L. Pedersen, and Patrik K.E. Magnusson

Table S1. General characteristics of study participants

	All	Gender		Zygosity[b]		
		Male	Female	MZ	SSDZ	OSDZ
Phenotypes						
n_{total}	10682	5074	5608	2499	4154	4029
Age[a]	64.89±8.08	65.45±7.99	64.38±8.13	63.67±7.49	65.45±8.47	65.06±7.95
$n_{pair\text{-}match}$	7740	3653	4087	2176	2886	2678
Age	65.03±7.75	65.58±7.69	64.54±7.77	63.54±7.31	65.63±8.09	65.60±7.56
Genotypes						
$n_{pheno\text{-}match}$	9606	4523	5083	1423	4154	4029
Age	65.04±8.15	65.48±8.05	64.65±8.22	63.76±7.62	65.45±8.47	65.06±7.95
$n_{unrelated}$	5779	2755	3024	1185	2316	2278
Age	64.91±8.33	65.51±8.21	64.37±8.40	63.96±7.71	65.38±8.69	64.93±8.22

[a] Age of each subgroup is described as mean ± standard deviation;

[b] Including monozygotic (MZ), same-sex dizygotic (SSDZ) and opposite-sex dizygotic (OSDZ) twins.

Table S2. Eighteen traits and units

Trait	Abbreviation	Unit
Total cholesterol	TC	mmol/L
High density lipoprotein	HDL	mmol/L
Low density lipoprotein	LDL	mmol/L
Apolipoprotein A1	ApoA1	g/L
Apolipoprotein B	ApoB	g/L
Triglyceride	TG	mmol/L
C-reactive Protein	CRP	mg/L
Glucose	Glu	mmol/L
Glycosylated hemoglobin A1c	HbA1c	%
Hemoglobin	HB	g/dL
Cystatin C	CysC	mg/L
Creatinine	Crea	μmol/L
Glomerular filtration rate	GFR	mL/min /1.73 m^2
Immunoglobulin A	IgA	g/L
Body mass index	BMI	kg/m^2
Weight	WT	kg
Waist circumference	WC	cm
Height		m

Table S3. Descriptive statistics of each trait

Trait	All	Gender[a]		Zygosity[b]		
		Male	Female	MZ	SSDZ	OSDZ
TC	5.77±1.12	5.52±1.10	6.00±1.09	5.77±1.12	5.79±1.12	5.76±1.13
n	10682	5074	5608	2499	4154	4029
HDL	1.41±0.42	1.24±0.34	1.56±0.42	1.40±0.43	1.42±0.42	1.39±0.42
n	10682	5074	5608	2499	4154	4029
LDL	3.76±0.99	3.65±0.98	3.86±0.99	3.78±0.98	3.76±0.99	3.74±0.99
n	10543	4985	5558	2467	4102	3974
ApoA1	1.64±0.30	1.53±0.26	1.75±0.30	1.64±0.31	1.64±0.30	1.64±0.30
n	10678	5073	5605	2497	4154	4027
ApoB	1.08±0.25	1.07±0.24	1.10±0.25	1.07±0.24	1.09±0.25	1.09±0.25
n	10678	5073	5605	2497	4154	4027
TG	1.37±0.82	1.44±0.92	1.30±0.72	1.33±0.78	1.36±0.83	1.39±0.83
n	10681	5073	5608	2499	4153	4029
CRP	3.13±6.61	3.29±7.95	3.00±5.10	3.05±6.80	3.14±6.84	3.18±6.24
n	10679	5074	5605	2498	4153	4028
Glu	5.59±1.22	5.76±1.33	5.44±1.08	5.53±1.09	5.60±1.25	5.63±1.26
n	10676	5073	5603	2499	4150	4027
HbA1c	4.82±0.68	4.85±0.74	4.78±0.61	4.77±0.60	4.82±0.67	4.84±0.72
n	10668	5069	5599	2498	4150	4020
HB	14.26±1.21	14.86±1.13	13.72±1.00	14.31±1.19	14.18±1.22	14.32±1.20
n	10664	5065	5599	2495	4148	4021
CysC	1.02±0.29	1.05±0.33	0.99±0.25	0.99±0.22	1.03±0.30	1.03±0.32
n	10405	4936	5469	2453	4020	3932
Crea	78.01±28.17	87.36±35.68	69.58±14.51	77.46±16.47	77.50±22.12	78.89±37.80
n	10405	4936	5469	2453	4020	3932
GFR	83.55±21.88	80.74±21.94	86.08±21.53	85.94±21.15	83.04±22.40	82.58±21.70
n	10404	4935	5469	2453	4020	3931
IgA	2.52±1.14	2.67±1.18	2.39±1.08	2.37±1.02	2.51±1.15	2.63±1.19
n	10621	5046	5575	2489	4139	3993
BMI	26.31±7.33	26.59±6.83	26.07±7.75	26.34±8.56	26.20±7.37	26.42±6.40
n	10562	5021	5541	2484	4110	3968
WT	74.94±13.81	81.77±12.33	68.75±12.05	75.11±13.95	74.18±13.59	75.62±13.91
n	10570	5027	5543	2487	4113	3970
WC	91.78±12.18	97.16±10.26	86.91±11.72	91.97±12.01	91.25±11.98	92.22±12.47
n	10556	5019	5537	2485	4106	3965
Height	1.69±0.10	1.76±0.09	1.63±0.08	1.70±0.10	1.69±0.10	1.70±0.10
n	10573	5026	5547	2486	4118	3969

[a] Sex-difference: TC, HDL, LDL, ApoA1, ApoB and GFR were significantly lower in males (P<0.05); TG, CRP, Glu, HbA1c, HB, CysC, Crea, IgA, BMI, weight, waist circumference and height were significantly higher in males (P<0.05);

[b] HDL, ApoB, TG, Glu, HbA1c, HB, CysC, Crea, GFR, IgA, weight, waist circumference and height were significantly different among monozygotic (MZ), same-sex dizygotic (SSDZ) and opposite-sex dizygotic (OSDZ) twins (P<0.05).

Table S4. Estimates and parameters in SNP-based GREML(d) analyses

Trait	GRM[a]	V(G)[b]	V(e)[c]	Vp[d]	V(G)/Vp[e]	logL[f]	logL0[g]	LRT[h]	df[i]	P[j]
TC	GREML	0.15	0.83	0.98	**0.15**	-2837.23	-2840.16	5.87	1	7.69E-3
(n=5779)	GREMLd	0.00	0.98	0.98	0.00	-2840.16	-2840.16	0.00	1	0.50
LDL	GREML	0.15	0.83	0.99	**0.16**	-2814.76	-2818.06	6.59	1	5.14E-3
(n=5704)	GREMLd	0.00	0.99	0.99	0.00	-2818.06	-2818.06	0.00	1	0.50
ApoB	GREML	0.14	0.85	0.99	**0.14**	-2861.98	-2864.65	5.33	1	0.01
(n=5776)	GREMLd	0.00	0.99	0.99	0.00	-2864.65	-2864.65	0.00	1	0.50
TG	GREML	0.31	0.68	0.99	**0.31**	-2841.81	-2856.49	29.35	1	3.02E-8
(n=5778)	GREMLd	0.28	0.71	0.99	0.28	-2851.29	-2856.49	10.40	1	6.29E-4
CRP	GREML	0.38	0.63	1.01	**0.37**	-2906.29	-2932.21	51.84	1	3.02E-13
(n=5777)	GREMLd	0.00	1.01	1.01	0.00	-2932.21	-2932.21	0.00	1	0.50
Glu	GREML	0.17	0.82	0.99	**0.17**	-2847.99	-2852.33	8.66	1	1.62E-3
(n=5775)	GREMLd	0.15	0.84	0.99	0.15	-2850.98	-2852.33	2.69	1	0.05
HbA1c	GREML	0.20	0.80	1.00	**0.20**	-2880.00	-2885.65	11.30	1	3.88E-4
(n=5770)	GREMLd	0.00	1.00	1.00	0.00	-2885.65	-2885.65	0.00	1	0.50
HB	GREML	0.21	0.80	1.01	**0.21**	-2921.46	-2927.49	12.07	1	2.56E-4
(n=5769)	GREMLd	0.00	1.01	1.01	0.00	-2927.49	-2927.49	0.00	1	0.50
CysC	GREML	0.28	0.74	1.02	**0.27**	-2877.14	-2886.99	19.71	1	4.51E-6
(n=5634)	GREMLd	0.05	0.97	1.02	0.05	-2886.82	-2886.99	0.34	1	0.28
Crea	GREML	0.18	0.83	1.01	**0.18**	-2848.49	-2853.10	9.21	1	1.20E-3
(n=5634)	GREMLd	0.01	1.01	1.01	0.01	-2853.09	-2853.10	0.00	1	0.48
GFR	GREML	0.33	0.69	1.02	**0.32**	-2869.81	-2883.10	26.59	1	1.26E-7
(n=5633)	GREMLd	0.04	0.99	1.02	0.03	-2883.03	-2883.10	0.14	1	0.35
BMI	GREML	0.20	0.78	0.98	**0.21**	-2792.70	-2799.26	13.11	1	1.47E-4
(n=5695)	GREMLd	0.02	0.96	0.98	0.02	-2799.22	-2799.26	0.08	1	0.39
WT	GREML	0.26	0.72	0.98	**0.26**	-2788.92	-2797.98	18.13	1	1.03E-5
(n=5699)	GREMLd	0.11	0.87	0.98	0.11	-2797.18	-2797.98	1.61	1	0.10
WC	GREML	0.16	0.81	0.97	**0.16**	-2763.68	-2767.08	6.81	1	4.55E-3
(n=5693)	GREMLd	0.19	0.78	0.97	0.19	-2764.65	-2767.08	4.86	1	0.01
HDL	GREML	0.24	0.77	1.01	**0.24**	-2911.87	-2919.78	15.83	1	3.47E-5
(n=5779)	GREMLd	0.01	1.00	1.01	0.01	-2919.77	-2919.78	0.03	1	0.44
ApoA1	GREML	0.17	0.84	1.00	**0.17**	-2899.83	-2903.44	7.23	1	3.58E-3
(n=5776)	GREMLd	0.09	0.92	1.00	0.09	-2902.92	-2903.44	1.05	1	0.15
IgA	GREML	0.25	0.77	1.02	**0.24**	-2916.66	-2925.59	17.88	1	1.18E-5
(n=5745)	GREMLd	0.00	1.01	1.02	0.00	-2925.59	-2925.59	0.00	1	0.49
Height	GREML	0.62	0.39	1.01	**0.62**	-2828.89	-2878.51	99.26	1	<0.01
(n=5701)	GREMLd	0.00	1.01	1.01	0.00	-2878.51	-2878.51	0.00	1	0.50

[a] Genomic-relatedness-matrix (additive or dominant); [b] Genetic variance (additive or dominant); [c] Non-shared environmental variance; [d] Phenotypic variance; [e] Additive or dominant genetic variance, significant dominance were labeled in red; [f] Log likelihood for the full model; [g] Log likelihood for the reduced model; [h] Likelihood ratio test; [i] Degree of freedom; [j] P value.

Table S5. Model fitting for classic twin-based structural equation model

Trait	Model[a]	EP[b]	minus 2LL[c]	Df[d]	AIC[e]	diff LL[f]	diff df[g]	P[h]
TC	Saturated	10	21583.58	7730	6123.58	NA	NA	NA
	ACE	4	21597.89	7736	6125.89	14.32	6	2.63E-02
	ADE	**4**	**21592.47**	**7736**	**6120.47**	**8.90**	**6**	**1.79E-01**
	AE	3	21597.89	7737	6123.89	14.32	7	4.58E-02
LDL	Saturated	10	21272.02	7629	6014.02	NA	NA	NA
	ACE	4	21287.48	7635	6017.48	15.46	6	1.70E-02
	ADE	**4**	**21279.49**	**7635**	**6009.49**	**7.47**	**6**	**2.79E-01**
	AE	3	21287.48	7636	6015.48	15.46	7	3.05E-02
ApoB	Saturated	10	21413.46	7728	5957.46	NA	NA	NA
	ACE	4	21424.26	7734	5956.26	10.81	6	9.45E-02
	ADE	**4**	**21421.19**	**7734**	**5953.19**	**7.74**	**6**	**2.58E-01**
	AE	3	21424.26	7735	5954.26	10.81	7	1.47E-01
TG	Saturated	10	21310.31	7729	5852.31	NA	NA	NA
	ACE	4	21324.27	7735	5854.27	13.96	6	3.01E-02
	ADE	**4**	**21320.95**	**7735**	**5850.95**	**10.64**	**6**	**1.00E-01**
	AE	3	21324.27	7736	5852.27	13.96	7	5.20E-02
CRP	Saturated	10	21587.85	7728	6131.85	NA	NA	NA
	ACE	4	21599.75	7734	6131.75	11.90	6	6.43E-02
	ADE	**4**	**21596.98**	**7734**	**6128.98**	**9.13**	**6**	**1.67E-01**
	AE	3	21599.75	7735	6129.75	11.90	7	1.04E-01
Glu	Saturated	10	21521.28	7726	6069.28	NA	NA	NA
	ACE	4	21551.71	7732	6087.71	30.43	6	3.26E-05
	ADE	**4**	**21537.24**	**7732**	**6073.24**	**15.96**	**6**	**1.40E-02**
	AE	3	21551.71	7733	6085.71	30.43	7	7.93E-05
HbA1c	Saturated	10	20836.70	7717	5402.70	NA	NA	NA
	ACE	4	20876.72	7723	5430.72	40.02	6	4.51E-07
	ADE	**4**	**20851.88**	**7723**	**5405.88**	**15.18**	**6**	**1.89E-02**
	AE	3	20876.72	7724	5428.72	40.02	7	1.25E-06
HB	Saturated	10	21246.00	7716	5814.00	NA	NA	NA
	ACE	4	21256.52	7722	5812.52	10.51	6	1.05E-01
	ADE	**4**	**21252.84**	**7722**	**5808.84**	**6.84**	**6**	**3.36E-01**
	AE	3	21256.52	7723	5810.52	10.51	7	1.61E-01
CysC	Saturated	10	20534.11	7524	5486.11	NA	NA	NA
	ACE	4	20552.35	7530	5492.35	18.23	6	5.67E-03
	ADE	**4**	**20546.69**	**7530**	**5486.69**	**12.57**	**6**	**5.04E-02**
	AE	3	20552.35	7531	5490.35	18.23	7	1.10E-02
Crea	Saturated	10	20720.13	7524	5672.13	NA	NA	NA
	ACE	4	20735.30	7530	5675.30	15.18	6	1.89E-02
	ADE	**4**	**20725.89**	**7530**	**5665.89**	**5.76**	**6**	**4.51E-01**
	AE	3	20735.30	7531	5673.30	15.18	7	3.38E-02
GFR	Saturated	10	20597.48	7524	5549.48	NA	NA	NA
	ACE	4	20612.62	7530	5552.62	15.14	6	1.92E-02
	ADE	**4**	**20605.68**	**7530**	**5545.69**	**8.20**	**6**	**2.24E-01**

	AE	3	20612.62	7531	5550.62	15.14	7	3.43E-02
BMI	Saturated	10	20876.61	7669	5538.61	NA	NA	NA
	ACE	4	20912.53	7675	5562.53	35.92	6	2.85E-06
	ADE	**4**	**20881.45**	**7675**	**5531.45**	**4.84**	**6**	**5.64E-01**
	AE	3	20912.53	7676	5560.53	35.92	7	7.50E-06
WT	Saturated	10	20721.53	7674	5373.54	NA	NA	NA
	ACE	4	20755.37	7680	5395.37	33.83	6	7.24E-06
	ADE	**4**	**20730.42**	**7680**	**5370.42**	**8.89**	**6**	**1.80E-01**
	AE	3	20755.37	7681	5393.37	33.83	7	1.85E-05
WC	Saturated	10	21035.29	7661	5713.29	NA	NA	NA
	ACE	4	21082.82	7667	5748.82	47.52	6	1.47E-08
	ADE	**4**	**21040.72**	**7667**	**5706.73**	**5.43**	**6**	**4.90E-01**
	AE	3	21082.82	7668	5746.82	47.52	7	4.41E-08
HDL	Saturated	10	20977.12	7730	5517.12	NA	NA	NA
	ACE	4	20988.98	7736	5516.98	11.86	6	6.51E-02
	ADE	4	20987.38	7736	5515.38	10.26	6	1.14E-01
	AE	**3**	**20988.98**	**7737**	**5514.98**	**11.86**	**7**	**1.05E-01**
ApoA1	Saturated	10	21035.10	7728	5579.10	NA	NA	NA
	ACE	4	21047.81	7734	5579.81	12.71	6	4.79E-02
	ADE	4	21048.01	7734	5580.01	12.91	6	4.44E-02
	AE	**3**	**21048.01**	**7735**	**5578.01**	**12.91**	**7**	**7.42E-02**
IgA	Saturated	10	21121.74	7681	5759.74	NA	NA	NA
	ACE	**4**	**21187.33**	**7687**	**5813.34**	**65.59**	**6**	**3.26E-12**
	ADE	4	21190.83	7687	5816.83	69.09	6	6.29E-13
	AE	3	21190.83	7688	5814.83	69.09	7	2.26E-12
Height	Saturated	10	19587.10	7675	4237.10	NA	NA	NA
	ACE	**4**	**19596.82**	**7681**	**4234.82**	**9.72**	**6**	**1.37E-01**
	ADE	4	19606.82	7681	4244.82	19.72	6	3.11E-03
	AE	3	19606.82	7682	4242.82	19.72	7	6.21E-03

[a] Each model is compared with saturated model, and the model in bold is the best fitted model according to Akaike information criterion;
[b] EP: number of estimated parameters;
[c] minus 2LL: minus 2 log likelihood;
[d] df: degree of freedom;
[e] AIC: Akaike information criterion;
[f] diff LL: difference of log likelihood;
[g] diff df: difference of df;
[h] P: P value.

Table S6. Intra-pair correlation and Falconer estimation for each trait

Trait	Intra-pair correlation[a]				Falconer Equation[b]		
	r_{MZ}	r_{DZ}	r_{SSDZ}	r_{OSDZ}	A	C	E
TC	**0.48**	**0.19**	**0.20**	**0.18**	**0.58**	**-0.10**	**0.52**
95% CI	(0.43-0.52)	(0.15-0.23)	(0.15-0.25)	(0.12-0.23)			
LDL	**0.46**	**0.18**	**0.19**	**0.16**	**0.56**	**-0.10**	**0.54**
95% CI	(0.41-0.51)	(0.14-0.21)	(0.14-0.24)	(0.11-0.22)			
ApoB	**0.52**	**0.23**	**0.26**	**0.20**	**0.58**	**-0.06**	**0.48**
95% CI	(0.48-0.57)	(0.20-0.27)	(0.21-0.31)	(0.15-0.25)			
TG	**0.55**	**0.24**	**0.27**	**0.21**	**0.62**	**-0.07**	**0.45**
95% CI	(0.51-0.59)	(0.21-0.28)	(0.23-0.32)	(0.16-0.26)			
CRP	**0.42**	**0.19**	**0.21**	**0.17**	**0.46**	**-0.04**	**0.58**
95% CI	(0.37-0.46)	(0.15-0.23)	(0.16-0.26)	(0.12-0.22)			
Glu	**0.51**	**0.20**	**0.23**	**0.17**	**0.62**	**-0.11**	**0.49**
95% CI	(0.46-0.55)	(0.16-0.24)	(0.18-0.28)	(0.11-0.22)			
HbA1c	**0.69**	**0.28**	**0.27**	**0.28**	**0.82**	**-0.13**	**0.31**
95% CI	(0.66-0.72)	(0.24-0.31)	(0.23-0.32)	(0.23-0.33)			
HB	**0.55**	**0.24**	**0.27**	**0.21**	**0.62**	**-0.07**	**0.45**
95% CI	(0.50-0.59)	(0.21-0.28)	(0.23-0.32)	(0.16-0.26)			
CysC	**0.57**	**0.26**	**0.27**	**0.26**	**0.62**	**-0.05**	**0.43**
95% CI	(0.53-0.61)	(0.23-0.30)	(0.22-0.32)	(0.20-0.31)			
Crea	**0.58**	**0.24**	**0.27**	**0.21**	**0.68**	**-0.10**	**0.42**
95% CI	(0.54-0.62)	(0.20-0.28)	(0.22-0.32)	(0.16-0.26)			
GFR	**0.57**	**0.24**	**0.24**	**0.24**	**0.66**	**-0.09**	**0.43**
95% CI	(0.53-0.61)	(0.21-0.28)	(0.19-0.29)	(0.19-0.29)			
BMI	**0.68**	**0.24**	**0.29**	**0.18**	**0.88**	**-0.20**	**0.32**
95% CI	(0.64-0.71)	(0.21-0.28)	(0.25-0.34)	(0.13-0.24)			
WT	**0.73**	**0.27**	**0.33**	**0.21**	**0.92**	**-0.19**	**0.27**
95% CI	(0.70-0.75)	(0.24-0.31)	(0.28-0.37)	(0.16-0.26)			
WC	**0.63**	**0.20**	**0.27**	**0.13**	**0.86**	**-0.23**	**0.37**
95% CI	(0.59-0.67)	(0.16-0.24)	(0.22-0.31)	(0.08-0.19)			
HDL	**0.67**	**0.31**	**0.34**	**0.28**	**0.72**	**-0.05**	**0.33**
95% CI	(0.64-0.70)	(0.28-0.34)	(0.29-0.38)	(0.23-0.33)			
ApoA1	**0.65**	**0.34**	**0.38**	**0.29**	**0.62**	**0.03**	**0.35**
95% CI	(0.61-0.68)	(0.30-0.37)	(0.33-0.42)	(0.24-0.34)			
IgA	**0.43**	**0.28**	**0.31**	**0.26**	**0.30**	**0.13**	**0.57**
95% CI	(0.38-0.48)	(0.25-0.32)	(0.26-0.35)	(0.21-0.31)			
Height	**0.87**	**0.48**	**0.49**	**0.46**	**0.78**	**0.09**	**0.13**
95% CI	(0.85-0.88)	(0.45-0.51)	(0.45-0.53)	(0.42-0.51)			

[a] All intra-pair correlations of each biomarker in each group were significant (P<0.0001);

[b] Falconer equations: $A=2(r_{MZ}-r_{DZ})$; $C=r_{MZ}-A$; $E=1-r_{MZ}$; A: additive genetic variance, C: shared environmental variance, E: non-shared/unique environmental variance.

Table S7. Estimates from classic and sex-limitation structural equation model

Traits		ACE model				ADE model				AE model	
		h^2	95% CI	c^2	95%CI	h^2	95% CI	d^2	95%CI	h^2	95% CI
Total cholesterol	All	0.45	(0.40,0.49)	0.00	(0.00,0.03)	**0.28**	**(0.13,0.43)**	**0.19**	**(0.03,0.36)**	0.45	(0.41,0.49)
	Male	0.42	(0.22,0.48)	0.00	(0.00,0.17)	0.43	(0.11,0.48)	0.00	(0.00,0.33)	**0.43**	**(0.36,0.48)**
	Female	0.51	(0.42,0.56)	0.00	(0.00,0.06)	0.28	(0.01,0.53)	0.25	(0.00,0.53)	**0.51**	**(0.45,0.56)**
LDL	All	0.43	(0.39,0,47)	0.00	(0.00,0.03)	**0.23**	**(0.08,0.38)**	**0.24**	**(0.07,0.41)**	0.43	(0.39,0.47)
	Male	0.38	(0.17,0.45)	0.00	(0.00,0.17)	**0.38**	**(0.04,0.45)**	**0.00**	**(0.00,0.36)**	0.38	(0.32,0.45)
	Female	0.52	(0.45,0.57)	0.00	(0.00,0.05)	**0.20**	**(0.00,0.46)**	**0.34**	**(0.07,0.59)**	0.52	(0.46,0.57)
Apolipoprotein B	All	0.51	(0.46,0.55)	0.00	(0.00,0.03)	**0.39**	**(0.25,0.53)**	**0.14**	**(0.00,0.30)**	0.51	(0.48,0.55)
	Male	0.40	(0.21,0.51)	0.05	(0.00,0.20)	0.46	(0.23,0.51)	0.00	(0.00,0.24)	**0.46**	**(0.40,0.51)**
	Female	0.61	(0.52,0.65)	0.00	(0.00,0.06)	0.42	(0.17,0.63)	0.19	(0.00,0.46)	**0.61**	**(0.56,0.65)**
Triglyceride	All	0.54	(0.49,0.58)	0.00	(0.00,0.03)	**0.42**	**(0.27,0.55)**	**0.14**	**(0.00,0.30)**	0.54	(0.51,0.58)
	Male	0.55	(0.40,0.60)	0.00	(0.00,0.12)	0.51	(0.22,0.60)	0.04	(0.00,0.35)	**0.55**	**(0.50,0.60)**
	Female	0.56	(0.46,0.61)	0.00	(0.00,0.08)	0.45	(0.20,0.61)	0.12	(0.00,0.39)	**0.56**	**(0.51,0.61)**
C-reactive protein	All	0.42	(0.36,0.46)	0.00	(0.00,0.04)	**0.30**	**(0.15,0.44)**	**0.14**	**(0.00,0.31)**	0.42	(0.38,0.46)
	Male	0.40	(0.27,0.46)	0.00	(0.00,0.10)	0.24	(0.00,0.45)	0.18	(0.00,0.46)	**0.40**	**(0.33,0.46)**
	Female	0.47	(0.31,0.53)	0.00	(0.00,0.00)	0.44	(0.18,0.53)	0.04	(0.00,0.32)	**0.47**	**(0.41,0.53)**
Glucose	All	0.50	(0.46,0.54)	0.00	(0.00,0.02)	**0.24**	**(0.09,0.38)**	**0.30**	**(0.15,0.46)**	0.50	(0.46,0.54)
	Male	0.54	(0.45,0.59)	0.00	(0.00,0.07)	**0.28**	**(0.00,0.56)**	**0.27**	**(0.00,0.58)**	0.54	(0.48,0.59)
	Female	0.52	(0.41,0.57)	0.00	(0.00,0.08)	**0.38**	**(0.13,0.56)**	**0.15**	**(0.00,0.42)**	0.52	(0.46,0.57)
HbA1c	All	0.70	(0.67,0.72)	0.00	(0.00,0.01)	**0.37**	**(0.24,0.51)**	**0.35**	**(0.21,0.49)**	0.70	(0.67,0.72)
	Male	0.72	(0.68,0.76)	0.00	(0.00,0.03)	**0.18**	**(0.00,0.47)**	**0.56**	**(0.26,0.76)**	0.72	(0.68,0.76)
	Female	0.69	(0.61,0.73)	0.00	(0.00,0.06)	**0.54**	**(0.30,0.72)**	**0.15**	**(0.00,0.40)**	0.69	(0.65,0.73)
Hemoglobin	All	0.54	(0.49,0.57)	0.00	(0.00,0.03)	**0.41**	**(0.26,0.55)**	**0.15**	**(0.00,0.30)**	0.54	(0.50,0.57)
	Male	0.53	(0.42,0.58)	0.00	(0.00,0.09)	0.38	(0.07,0.58)	0.16	(0.00,0.49)	**0.53**	**(0.48,0.58)**
	Female	0.54	(0.39,0.63)	0.04	(0.00,0.16)	0.58	(0.40,0.63)	0.00	(0.00,0.19)	**0.58**	**(0.53,0.63)**
Cystatin C	All	0.59	(0.54,0.62)	0.00	(0.00,0.03)	**0.42**	**(0.28,0.56)**	**0.18**	**(0.03,0.34)**	0.59	(0.55,0.62)
	Male	0.53	(0.45,0.58)	0.00	(0.00,0.06)	**0.21**	**(0.00,0.51)**	**0.34**	**(0.03,0.59)**	0.53	(0.47,0.58)

	Female	0.68	(0.58,0.71)	0.00	(0.00,0.08)	**0.59**	**(0.34,0.71)**	**0.09**	**(0.00,0.35)**	0.68	(0.63,0.71)
Creatinine	All	0.57	(0.53,0.60)	0.00	(0.00,0.02)	**0.35**	**(0.21,0.50)**	**0.24**	**(0.09,0.40)**	0.57	(0.53,0.60)
	Male	0.56	(0.49,0.61)	0.00	(0.00,0.04)	**0.09**	**(0.00,0.39)**	**0.50**	**(0.18,0.63)**	0.56	(0.50,0.61)
	Female	0.49	(0.35,0.64)	0.12	(0.00,0.24)	**0.63**	**(0.52,0.67)**	**0.00**	**(0.00,0.11)**	0.63	(0.58,0.67)
GFR	All	0.56	(0.52,0.60)	0.00	(0.00,0.03)	**0.38**	**(0.23,0.52)**	**0.21**	**(0.05,0.36)**	0.56	(0.53,0.60)
	Male	0.51	(0.44,0.57)	0.00	(0.00,0.05)	**0.08**	**(0.00,0.38)**	**0.46**	**(0.14,0.59)**	0.51	(0.45,0.57)
	Female	0.63	(0.50,0.67)	0.00	(0.00,0.11)	**0.61**	**(0.35,0.67)**	**0.03**	**(0.00,0.29)**	0.63	(0.59,0.67)
Body mass index	All	0.65	(0.62,0.68)	0.00	(0.00,0.01)	**0.28**	**(0.13,0.42)**	**0.41**	**(0.26,0.56)**	0.65	(0.62,0.68)
	Male	0.66	(0.59,0.70)	0.00	(0.00,0.06)	**0.37**	**(0.06,0.66)**	**0.30**	**(0.01,0.62)**	0.66	(0.61,0.70)
	Female	0.68	(0.60,0.72)	0.00	(0.00,0.07)	**0.56**	**(0.32,0.71)**	**0.12**	**(0.00,0.38)**	0.68	(0.64,0.72)
Weight	All	0.70	(0.68,0.73)	0.00	(0.00,0.01)	**0.37**	**(0.23,0.51)**	**0.35**	**(0.21,0.50)**	0.70	(0.68,0.73)
	Male	0.73	(0.63,0.76)	0.00	(0.00,0.09)	0.63	(0.34,0.76)	0.10	(0.00,0.40)	**0.73**	**(0.70,0.76)**
	Female	0.71	(0.62,0.74)	0.00	(0.00,0.08)	0.61	(0.36,0.74)	0.10	(0.00,0.36)	**0.71**	**(0.67,0.74)**
Waist circumference	All	0.60	(0.56,0.63)	0.00	(0.00,0.01)	**0.15**	**(0.01,0.29)**	**0.49**	**(0.34,0.65)**	0.60	(0.56,0.63)
	Male	0.63	(0.56,0.68)	0.00	(0.00,0.06)	**0.32**	**(0.01,0.62)**	**0.32**	**(0.02,0.65)**	0.63	(0.59,0.68)
	Female	0.63	(0.54,0.67)	0.00	(0.00,0.08)	**0.51**	**(0.25,0.67)**	**0.13**	**(0.00,0.40)**	0.63	(0.58,0.67)
HDL	All	0.66	(0.61,0.69)	0.00	(0.00,0.04)	0.58	(0.44,0.68)	0.09	(0.00,0.23)	**0.66**	**(0.63,0.69)**
	Male	0.60	(0.44,0.68)	0.05	(0.00,0.18)	0.65	(0.44,0.69)	0.00	(0.00,0.21)	**0.65**	**(0.60,0.69)**
	Female	0.67	(0.56,0.71)	0.00	(0.00,0.10)	0.63	(0.39,0.71)	0.04	(0.00,0.29)	**0.67**	**(0.63,0.71)**
Apolipoprotein A1	All	0.64	(0.55,0.68)	0.02	(0.00,0.08)	0.66	(0.55,0.68)	0.00	(0.00,0.11)	**0.66**	**(0.63,0.68)**
	Male	0.55	(0.40,0.67)	0.07	(0.00,0.20)	0.63	(0.46,0.67)	0.00	(0.00,0.17)	**0.63**	**(0.59,0.67)**
	Female	0.60	(0.47,0.71)	0.07	(0.00,0.18)	0.68	(0.55,0.71)	0.00	(0.00,0.13)	**0.68**	**(0.64,0.71)**
Immunoglobulin A	All	**0.40**	**(0.29,0.51)**	**0.07**	**(0.00,0.15)**	0.50	(0.43,0.54)	0.00	(0.00,0.07)	0.50	(0.46,0.54)
	Male	**0.13**	**(0.00,0.32)**	**0.26**	**(0.11,0.39)**	0.43	(0.33,0.48)	0.00	(0.00,0.09)	0.43	(0.37,0.48)
	Female	**0.59**	**(0.47,0.64)**	**0.00**	**(0.00,0.10)**	0.52	(0.27,0.63)	0.07	(0.00,0.34)	0.59	(0.53,0.64)
Height	All	**0.77**	**(0.71,0.83)**	**0.09**	**(0.04,0.15)**	0.87	(0.83,0.88)	0.00	(0.00,0.03)	0.87	(0.85,0.88)
	Male	**0.72**	**(0.62,0.84)**	**0.14**	**(0.03,0.24)**	0.86	(0.77,0.88)	0.00	(0.00,0.09)	0.86	(0.85,0.88)
	Female	**0.75**	**(0.65,0.86)**	**0.12**	**(0.01,0.21)**	0.87	(0.78,0.89)	0.00	(0.00,0.09)	0.87	(0.85,0.89)

The values in bold black are the estimates from the best fitting model.

Table S8. Model fitting for sex-limitation structural equation model

Trait	Model[a]	EP[b]	minus 2LL[c]	Df[d]	AIC[e]	diff LL[f]	diff df[g]	P[h]
	ACE	17	21580.91	7723	6134.914	NA	NA	NA
TC	ADE	17	21577.63	7723	6131.626	-3.29	0	1
	AE	**15**	**21580.92**	**7725**	**6130.918**	**0.00**	**2**	**1**
	ACE	17	21263.85	7622	6019.847	NA	NA	NA
LDL	**ADE**	**17**	**21257.72**	**7622**	**6013.719**	**-6.13**	**0**	**1**
	AE	15	21263.85	7624	6015.847	0.00	2	1
	ACE	17	21396.60	7721	5954.595	NA	NA	NA
ApoB	ADE	17	21394.70	7721	5952.696	-1.90	0	1
	AE	**15**	**21396.95**	**7723**	**5950.949**	**0.35**	**2**	**0.84**
	ACE	17	21055.15	7722	5611.154	NA	NA	NA
TG	ADE	17	21054.26	7722	5610.256	-0.90	0	1
	AE	**15**	**21055.15**	**7724**	**5607.154**	**0.00**	**2**	**1**
	ACE	17	21488.40	7721	6046.404	NA	NA	NA
CRP	ADE	17	21487.23	7721	6045.231	-1.17	0	1
	AE	**15**	**21488.40**	**7723**	**6042.404**	**0.00**	**2**	**1**
	ACE	17	21345.34	7719	5907.342	NA	NA	NA
Glu	**ADE**	**17**	**21341.11**	**7719**	**5903.110**	**-4.23**	**0**	**1**
	AE	15	21345.34	7721	5903.342	0.00	2	1
	ACE	17	20747.08	7710	5327.083	NA	NA	NA
HbA1c	**ADE**	**17**	**20730.82**	**7710**	**5310.823**	**-16.26**	**0**	**1**
	AE	15	20747.08	7712	5323.083	0.00	2	1
	ACE	17	21190.62	7709	5772.621	NA	NA	NA
HB	ADE	17	21189.98	7709	5771.978	-0.64	0	1
	AE	**15**	**21190.97**	**7711**	**5768.968**	**0.35**	**2**	**0.84**
	ACE	17	20468.89	7517	5434.887	NA	NA	NA
CysC	**ADE**	**17**	**20463.73**	**7517**	**5429.728**	**-5.16**	**0**	**1**
	AE	15	20468.89	7519	5430.887	0.00	2	1
	ACE	17	20505.86	7517	5471.856	NA	NA	NA
Crea	**ADE**	**17**	**20499.42**	**7517**	**5465.422**	**-6.43**	**0**	**1**
	AE	15	20509.35	7519	5471.350	3.49	2	0.17
	ACE	17	20580.29	7517	5546.288	NA	NA	NA
GFR	**ADE**	**17**	**20571.89**	**7517**	**5537.887**	**-8.40**	**0**	**1**
	AE	15	20580.29	7519	5542.288	0.00	2	1
	ACE	17	20782.26	7662	5458.262	NA	NA	NA
BMI	**ADE**	**17**	**20777.22**	**7662**	**5453.224**	**-5.04**	**0**	**1**
	AE	15	20782.26	7664	5454.262	0.00	2	1
	ACE	17	20697.38	7667	5363.382	NA	NA	NA
WT	ADE	17	20696.19	7667	5362.193	-1.19	0	1
	AE	**15**	**20697.38**	**7669**	**5359.382**	**0.00**	**2**	**1**
	ACE	17	20979.99	7654	5671.987	NA	NA	NA
WC	**ADE**	**17**	**20974.49**	**7654**	**5666.491**	**-5.50**	**0**	**1**
	AE	15	20979.99	7656	5667.987	0	2	1

HDL	ACE	17	20813.13	7723	5367.129	NA	NA	NA
	ADE	17	20813.46	7723	5367.459	0.33	0	0
	AE	**15**	**20813.57**	**7725**	**5363.575**	**0.45**	**2**	**0.80**
ApoA1	ACE	17	20986.29	7721	5544.293	NA	NA	NA
	ADE	17	20988.84	7721	5546.838	2.55	0	0
	AE	**15**	**20988.84**	**7723**	**5542.838**	**2.55**	**2**	**0.28**
IgA	**ACE**	**17**	**21088.02**	**7674**	**5740.018**	**NA**	**NA**	**NA**
	ADE	17	21098.33	7674	5750.325	10.31	0	0
	AE	15	21098.64	7676	5746.638	10.62	2	0.00
Height	**ACE**	**17**	**19551.80**	**7668**	**4215.802**	**NA**	**NA**	**NA**
	ADE	17	19562.34	7668	4226.342	10.54	0	0
	AE	15	19562.34	7670	4222.342	10.54	2	0.01

[a] The model in bold is the best fitted model according to Akaike information criterion;
[b] EP: number of estimated parameters;
[c] minus 2LL: minus 2 log likelihood;
[d] df: degree of freedom;
[e] AIC: Akaike information criterion;
[f] diff LL: difference of log likelihood;
[g] diff df: difference of df;
[h] P: P value.

Table S9. Distributions of the contact frequency and separation age by zygosity and sex

	MZ			SSDZ			OSDZ
	All	**Male**	**Female**	**All**	**Male**	**Female**	
Contact freqency							
n_{pair}	1066	549	517	1424	596	828	1325
Mean	3.03	3.04	3.03	2.71	2.64	2.76	2.45
Standard Deviation	0.82	0.82	0.81	0.82	0.81	0.82	0.69
Separation age							
n_{pair}	1027	518	509	1393	579	814	1302
Mean	19.80	20.29	19.30	18.55	19.30	18.02	18.25
Standard Deviation	3.43	3.59	3.20	3.59	3.76	3.37	3.75

Contact frequency was defined by the replies to questions about how often the participants usually met with their co-twin, and it was divided into four levels: 1-less than once a year; 2-on a yearly basis; 3-on a monthly basis; or 4-on a weekly basis; the measure here is used as a continuous variable. Contact frequency and separation age are significantly different between monozygotic (MZ) and same-sex dizygotic (SSDZ) twins ($P<0.001$), MZ and opposite-sex dizygotic(OSDZ) twins ($P<0.001$), SSDZ and OSDZ ($P<0.05$); Contact frequency is not significantly different between males and females in MZ ($P=0.94$), but significantly in SSDZ ($P=0.007$); separation ages are significantly different between males and females both in MZ and in SSDZ ($P<0.001$); Contact frequency and separation age are both significantly different between males of MZ and SSDZ($P<0.001$), and females of MZ and SSDZ ($P<0.001$).

Table S10. Intra-pair correlation stratified by level of shared environment in monozygotic twins

Trait	1) Contact frequency[a]				2) Separation age[b]			
	r_{Low}	n_{pair}	r_{High}	n_{pair}	r_{Low}	n_{pair}	r_{High}	n_{pair}
Total cholesterol	0.461	619	0.482	447	0.497	530	0.440	497
High density lipoprotein	0.657	619	0.690	447	0.658	530	0.690	497
Low density lipoprotein	0.442	610	0.479	434	0.484	523	0.430	484
Apolipoprotein A1	0.625	618	0.677	447	0.647	529	0.663	497
Apolipoprotein B	0.504	618	0.542	447	0.521	529	0.521	497
Triglyceride	0.547	619	0.575	447	0.484	530	0.613	497
C-reactive protein	0.384	619	0.471	446	0.458	529	0.389	497
Glucose	0.512	619	0.501	447	0.487	530	0.516	497
Glycosylated hemoglobin A1c	0.644	618	0.750	447	0.657	529	0.737	497
Hemoglobin	0.543	616	0.539	447	0.537	528	0.559	496
Cystatin C	0.530	600	0.635	429	0.546	514	0.596	479
Creatinine	0.541	600	0.630	429	0.567	514	0.606	479
Glomerular filtration rate	0.533	600	0.623	429	0.544	514	0.588	479
Immunoglobulin A	0.447	614	0.402	444	0.474	529	0.395	491
Body mass index	0.641	614	0.727	447	0.647	527	0.708	495
Weight	0.676	617	0.794	447	0.700	529	0.753	496
Waist circumference	0.602	619	0.689	444	0.620	528	0.645	496
Height	0.846	616	0.882	447	0.820	528	0.906	496

[a] Low contact frequency was defined as ≤3 (on the monthly basis), high contact frequency was defined as >3; contact frequency in low group (mean=2.43, standard deviation=0.48) was significantly lower than high group (mean=3.88, standard deviation=0.22), P<0.0001; r is the intra-pair correlation coefficient, all of them were significant (P< 0.0001);

[b] Low separation age was defined as ≤19.5 years old (median age), while high separation age was defined as > 19.5 years old; separation age in low group (mean=17.46, standard deviation=2.23) was significantly younger than high group (mean=22.30, standard deviation=2.63), P<0.0001; r is the intra-pair correlation coefficient, all of them were significant (P<0.0001).