THE UNIVERSITY
of EDINBURGH

The Architecture of Human Complex Trait Variation

Charley XIA

(registered as Xiachi XIN)

Doctor of Philosophy (PhD)

THE UNIVERSITY *of* EDINBURGH

2018

Declaration

I declare that this thesis was written by myself and the research reported within this thesis is my own work, under supervision of Dr. Pau Navarro and Professor. Chris Haley. My contribution to the work in the article attached in the thesis is detailed under "Publications".

This work has not been submitted for the award of any degree at any institution.

Signature:_____

Date:_____

Acknowledgements

I would like to thank the following persons for their assistance. Without their help, this thesis would never have been accomplished.

Firstly, I would like to express my greatest appreciation to my primary supervisor Prof. Chris Haley. He recruited me into his group, generously provided me with stipend and patiently handled all my queries. He is a great mentor who guides me through my PhD study.

Secondly, I would like to acknowledge the help provided by my second supervisor Dr. Pau Navarro. Advice given by her has been a great help in my work, study and livelihood through this journey.

Subsequently, I am grateful for the assistance and encouragement given by current group members Carmen Amador, Yanni Zeng and Andrew Bretherick and by previous group members Masoud Shirali, Athina Spiliopoulou and Mairead Bermingham. I also appreciate the efforts of my thesis committee members, including Dr. Veronique Vitart, Prof. Colin Semple and my supervisors.

Afterwards, I would like to express my deep gratitude to Medical Research Council (UK) and the University of Edinburgh for awarding my studentship and offering me funding and I would like to thank Generation Scotland committee for the collection of the data.

Finally, I thank my wife and mother for their love, encouragement and support during this period.

Publications

The following publication has resulted as a direct outcome of the research described in this thesis. I am the first author of this publication who designed the study, analysed the data and composed the manuscript. The publication is attached in the main text in Chapter 2 and its supplementary information is attached in Appendix.

Xia C, Amador C, Huffman J, et al. (2016) Pedigree- and SNP-Associated Genetics and Recent Environment are the Major Contributors to Anthropometric and Cardiometabolic Trait Variation. *PLOS Genetics* 12(2): e1005804.

The following publication has resulted from research associated with this thesis. I am one of the co-first authors of this publication who designed the study, analysed the data and composed a part of the manuscript. Some results from this publication are extracted and described in the discussion section in Chapter 2 as further support for my findings. The publication is attached in the Appendix.

Hill WD, Arslan RC, **Xia C**, et al. (2018 Genomic analysis of family data reveals additional genetic effects on intelligence and personality. *Molecular Psychiatry*:ePrint.

The following publications have resulted from research associated with this thesis. I am one of the second authors of these publications who contributed in methodology and preparing the materials for analysis. Some results from the first publication listed below have been extracted and described in the discussion section in Chapter 2 as further support for my findings; whereas the conclusion from the second publication listed below is introduced in the discussion section in Chapter 2 as an application of my research. The publications are attached in the Appendix.

Zeng Y, Navarro P, **Xia C**, et al. (2016) Shared Genetics and Couple-Associated Environment Are Major Contributors to the Risk of Both Clinical and Self-Declared Depression. *EBioMedicine* 14:161 – 167.

Amador C, **Xia C**, Nagy R, et al. (2017) Regional variation in health is predominantly driven by lifestyle rather than genetics. *Nature Communications* 8: 801.

The following unpublished manuscript has resulted as a direct outcome of the research described in this thesis. I am one of the co-first authors of this study who contributed to the analysis and composed a part of the manuscript. The results from this study were presented in the main text in Chapter 3 and the draft manuscript is attached in Appendix.

Canela-Xandri O, **Xia C**, Rawlik K, et al. (2017) New Tools for Genome-Wide Association Studies on A Population Level Scale.

Lay Summary

A complex trait or disease is one for which the differences between people are influenced by both genes and environment (lifestyle, diet, etc.), along with their interactions. Uncovering which factors determine variation in a trait (its "architecture") will help us find ways to treat or avoid disease. This thesis explores the architecture of human complex traits related to body shape and cardiovascular health, such as height, body mass index and blood pressure, using data from ~20k individuals of recent Scottish descent from a study called Generation Scotland: Scottish Family Health Study.

I conducted a statistical analysis to quantify the effects of different origins on trait variation, e.g. how much an effect contributes to the trait differences between people. I explored the relative influence of genes that occur commonly in the general population compared to that of rare mutations running in families, and the influence of sharing a common environment, either with all family members, or your partner, or your siblings. Genetics was the most important factor investigated, explaining ~45% of the differences between individuals on average for body shape and cardiovascular health. Common genetic variations affect the trait to the same extent as rare mutations shared within families. The common environment shared by all family members makes a limited contribution to trait variation; whereas common couple environment and common sibling environment could explain another 11% and 5% of the trait differences on average.

Genome-Wide Association Studies (GWAS) aim to identify causal genes and genetic variants of a trait. The ability of GWAS to detect genetic causal factors is enhanced the more we know about the trait architecture. Knowledge of trait architecture could also be used for disease risk prediction, e.g. to predict a person's tendency to be overweight. Since common couple and sibling environment were found to explain a substantial amount of the trait differences between individuals, I implemented them into GWAS analyses and prediction models and concluded that taking them into account helps us find more causal genes and improves the ability to predict, compared to the common GWAS and prediction methods using only genetic information.

In my thesis, the high similarity between partners was considered as environmental effects whilst it could also be due to non-random mate choice. For example, when tall people marry tall people, becoming more alike due to a shared environment is clearly not an explanation of why they are both tall. I explored what is the effect of this 'assortative mating' (people partnering someone of a similar phenotype) on trait architecture and developed a novel way to estimate the genetic contribution on trait differences using the similarity between relatives and in-laws.

In summary, my method offers a novel way of using information on family members (lifestyle, diet, etc), over and above the fact we know little about the real environmental causal factors affecting the traits. I discovered that sibling environment and couple effects, including shared couple environment and assortative mating, significantly contribute to the variation observed in body shape and cardiovascular health between individuals in our Scottish samples. Accounting for these effects is beneficial for disease prediction and to uncover which genes affect our health.

Abstract

A complex trait is a trait or disease that is controlled by both genetic and environmental factors, along with their interactions. Trait architecture encompasses the genetic variants and environmental causes of variation in the trait or disease, their effects on the trait or disease and the mechanism by which these factors interact at molecular and organism levels. It is important to understand trait architecture both from a biological viewpoint and a health perspective. In this thesis, I laid emphasis on exploring the influence of familial environmental factors on complex trait architecture alongside the genetic components. I performed a variety of studies to explore the architecture of anthropometric and cardio-metabolic traits, such as height, body mass index, high density lipoprotein content of blood and blood pressure, using a cohort of 20,000 individuals of recent Scottish descent and their phenotype measurements, Single Nucleotide Polymorphism (SNP) data and genealogical information.

I extended a method of variance component analysis that could simultaneously estimate SNP-associated heritability and total heritability whilst considering familial environmental effects shared among siblings, couples and nuclear family members. I found that most missing heritability could be explained by including closely related individuals in the analysis and accounting for these close relationships; and that, on top of genetics, couple and sibling environmental effects are additional significant contributors to the complex trait variation investigated.

Subsequently, I accounted for couple and sibling environmental effects in Genome-Wide Association Study (GWAS) and prediction models. Results demonstrated that by adding additional couple and sibling information, both GWAS performance and prediction accuracy were boosted for most traits investigated, especially for traits related to obesity.

Since couple environmental effects as modelled in my study might, in fact, reflect the combined effect of assortative mating and shared couple environment, I explored further the dissection of couple effects according to their origin. I extended assortative mating theory by deriving the expected resemblance between an individual and in-laws of his first-degree relatives. Using the expected resemblance derived, I developed

a novel pedigree study which could jointly estimate the heritability and the degree of assortative mating.

I have shown in this thesis that, for anthropometric and cardio-metabolic traits, environmental factors shared by siblings and couples seem to have important effects on trait variation and that appropriate modelling of such effects may improve the outcome of genetic analyses and our understanding of the causes of trait variation. My thesis also points out that future studies on exploring trait architecture should not be limited to genetics because environment, as well as mate choice, might be a major contributor to trait variation, although trait architecture varies from trait to trait.

Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Overview

A complex trait (or disease) is a trait (or disease) contributed by both genes and environment, together with interactions, if any. The thesis, entitled 'Exploring the Trait Architecture of Human Complex Trait', records my projects of studying the trait architecture of complex traits in humans. Prior to investigating human complex trait architecture, initially I need to comprehend what trait architecture is and what should be involved in trait architecture studies.

Therefore, to answer the question of what trait architecture is and how I propose to study it:

First, I broadly reviewed the relevant history, including important genetic and genomic discoveries and the development of genetic techniques and research methods;

Second, based on the knowledge learned from the history, I address my understanding of trait architecture. Afterwards, I have classified the trait architecture studies into a number of categories and I summarise the current status and problems for each category as well as the common problems remaining for the whole trait architecture research field;

Finally, I outline the aims of this project and I summarise the outline of my thesis.

Note, the review in Chapter 1 does not contain any details of the methodologies, equations or software used to conduct a specified study. On the contrary, it serves as background introduction, providing an overall view of the field of trait architecture study nowadays. I provide a summary of the appropriate methodology at the beginning of corresponding chapters.

1.2 Relevant History: Short Review

1.2.1 Early Times

1.2.1.1 The Beginning of Modern Genetics: 1850s-1910s

Unlike some other scientific disciplines such as mathematics, chemistry and physics, genetics is a relatively young subject with only about 150 years of history. The father of modern genetics is Gregor Mendel, who conducted pea plant experiments between 1856 and 1863 and established Mendel's laws of inheritance, known as the principles of dominance, independent assortment and segregation [1,2]. However, his work was not widely recognised in his lifetime until rediscovered by three other researchers in 1890s [2], including Hugo de Vries who proposed the concept of "pangenes", which was named after Darwin's "pangenesis" theory [2-4] and was abbreviated to the still-in-use-name "gene" by Wilhelm Johannsen in 1909 [2,5], for describing the smallest hereditary unit of specific traits in organisms [2,6]. Shortly after this, Thomas H. Morgan, who initiated molecular genetics, together with his student Alfred Sturtevant found sex-linkage and recombination frequency of genes by studying *Drosophila melanogaster* between 1910 and 1911 [2,7].

By then, the underlying structure and functions of the gene had not yet been revealed, a gene was no more than a postulated concept. However, the basic roles of how they are inherited were detected and the relationship between genotypes and phenotypes for simple traits (monogenic and oligogenic traits) that follow Mendelian rules was deduced.

1.2.1.2 The Development of Statistical Genetics: 1910s-1920s

Later on, in 1915 John Haldane first demonstrated the concept of genetic linkage in his study of mice [8]. In 1918, Ronald A. Fisher put forward a variance component model to estimate the trait variance attributable to genetics and to the environment by using the resemblance between different types of relative [9]. The ratio of genetic to total variance was subsequently termed the heritability [10]. Sewall Wright developed the theory of inbreeding, F-statistics, genetic drift and the path analysis method and

introduced different mating systems (such as random matting and assortative mating) in 1920s [11-13]. Together these three scientists developed the foundation of evolutionary, population and quantitative genetics.

With the guidance of the pioneers' tremendous work, scientists, regardless of knowing anything about gene function, were capable of calculating the linkage between two genes, estimating to what extent the trait or disease was heritable (low or high heritability) using related individuals and researching into the influence of historical events and human behaviours, such as inbreeding, assortative mating and drift, on genetic inheritance and its diversity, both of which give us broad insights into trait genetic architecture. Research methods invented back then, like pedigree studies of heritability (twin studies and parent-offspring regression) and F-statistics, are still widely in use in population and evolutionary genetics nowadays [14-16].

1.2.2 The Second Half of the 20th Century

1.2.2.1 DNA Era – The Discovery of Gene: 1940s-1970s

Scientists developed a more and more profound understanding of the essence, structure and function of genes over the period after the 1940s, from the Avery–MacLeod–McCarty experiment demonstrating deoxyribonucleic acid (DNA) as the genetic material in 1944 [17] to the discovery of the double helix structure of DNA by James Watson and Francis Crick in 1953 [18]; from the identification of the function of messenger ribonucleic acid (mRNA) by Sydney Brenner, Francois Jacob and Matthew Meselson in 1961 [19] to the complete decoding of triplet genetic codons by Marshall Nirenberg, Har G. Khorana, Sydney Brenner and Francis Crick by 1965 [20-22]; from the first gene sequence of bacteriophage MS2 coat protein gene by Walter Fiers in 1972 [23] to the finding of alternative splicing by Phillip Sharp and Richard Roberts in 1977 [24,25].

1.2.2.2 Linkage Mapping Era – The Usage of Genetic Marker: 1970s-2000s

With the development of genetic theories (1.2.1 Early Time) and the knowledge of genes (1.2.2.1 DNA Era), scientists were able to identify causal loci of a trait by linkage mapping studies using genetic markers.

The first generation of DNA-based genetic markers were restriction fragment length polymorphisms (RFLP), which was found in 1974 [26] and first used for linkage mapping in 1988 [27,28] due to RFLP mapping theory proposed by David Botstein in 1980 [29]. By mapping in experimental crosses of tomato, the first complete RFLP linkage mapping study identified 6 QTLs associated with fruit mass [28].

The second generation of genetic markers is variable number tandem repeat (VNTR), classified as microsatellites (i.e. a repeating unit which is less than < 5bp), also known as simple sequence repeat (SSR) and short tandem repeat (STR), and minisatellites (i.e. a repeating unit which is more than > 5bp). VNRTs, identified and proposed for mapping by Yusuke Nakamura in 1987 [30], were frequently used in genetics studies in 1990s [31,32] and early 2000s [33].

1.2.3 Recent Time

1.2.3.1 Genome Projects in Humans – The Understanding of Human Genome: 1990s-Now

The human genome is made up of both coding DNA (genes) and non-coding DNA (which do not encode proteins). To reach a better understanding of the genome structure and exploring the role of coding and non-coding DNA as well as their variations, scientists over the world joined forces and decided to research the genome of human as a consortium.

The first and most important genome project in humans was the Human Genome Project (HGP), which was proposed in late 1980s, initiated in 1990 and completed in 2003, aiming to sequence the whole human genome [34]. HGP provides us with vital information about the length of genome, the number of genes and proteins, and the location of coding, non-coding and pseudogene regions, etc., which greatly improved

our genomic knowledge. The aligned sequence results are often used as reference genome for blast, primers and chip design, imputation and annotation. In build 35 (2003's version), the coverage was 99% with accuracy of 99.999% and there were 341 gaps [34]. Nowadays, scientists keep trying to narrow down the gaps by sequencing or re-sequencing. The human reference genome is maintained and updated by the Genome Reference Consortium (GRC) and the current build is GRCh38.p8 (by August 2016) [35]. All this information is widely used in genetic and genomic research today.

The HGP maps genetic variations, such as single nucleotide polymorphisms (SNPs) and structural variations, on the genome, which enriches the choices of genetic markers. Subsequent projects, such as the International HapMap Project (HapMap, from 2002 to 2007) [36,37], the Human Variome Project (HVP, from 2006 to now) [38,39], the 1000 Genome Project (1000Genome, from 2008 to 2015) [40,41] and the 100,000 Genome Project (from 2012 to now) [42], all focus on identifying various genetic variations in the human genome.

1.2.3.2 Limitation of Linkage Study and the Development of SNP Array

Linkage studies in the 20th century were limited by the following. A), restricted abundance of RFLPs and VNRTs and the cumbersome processes of polymerase chain reaction (PCR) and gel electrophoresis. B), linkage study requires informative pedigree(s), and making it hard to recruit numerous people. C) linkage study only utilises within family information and hence the information between unrelated individuals is squandered. D), a linkage map is a genetic map which indicates the genetic location between a marker and a gene based on recombination frequency but not the absolute location on the genome. Even if a gene could be perfectly pinned down at a 1-2 centi-Morgan (cM) gap, the length of the gap is equivalent to 1-2 million base pairs (bp), and yet, it is still difficult to identify the candidate gene within this region. E), linkage study works well on the detection of causal genetic variants for simple traits (monogenic and oligogenic traits following Mendelian rules) but less successfully on complex traits attributed by polygenes and environmental effects, along with interaction if any. Although most variants showing Mendelian inheritance (4,773 according to online database, by August 2016) are studied, known and

published on Online Mendelian Inheritance in Man (OMIM) [43,44], many common diseases, such as type-II diabetes, obesity, bipolar, cardiovascular diseases and various types of cancer and tumour, are complex diseases, the studies of which were unsatisfactory by linkage analysis and the knowledge of causal factors for these was limited.

However, the appearance of microarray technology and the popularisation of association study largely eliminated these deficiencies.

Microarrays can be used to genotype thousands of markers simultaneously through an automatic machine, which greatly saves time and effort. The most common microarray is the SNP chip.

By 2001, there were ~1.4 million identified SNPs in human genome, mostly (~95%) identified by the HGP and The SNP Consortium (TSC) [45]. By August 2016, according to the online Single Nucleotide Polymorphism Database (dbSNP) [46], the number has increased tremendously to ~165 million owing to the follow-up genomic studies such as HapMAP and 1000Genome.

Due to the abundance of SNPs across the genome (1 per 18 bp on average), they have been widely used as markers in genetic analysis and for genotyping since late 1990s [45,47], e.g. the largest cohort UK biobank (over 500,000 people and 73 million imputed SNPs since 2006) [48]. The genomic relationship between individuals of unknown relationship can be derived by using dense markers such as SNPs, owing to Paul M. VanRaden [49,50], thus easing recruitment. Besides, the genomic relationship between nominally unrelated individuals estimated using SNP chip can be used for heritability study [51,52].

1.2.3.3 Association Mapping Era – The Golden Age of Genetic Studies: 2000s-Now

With the help of several human genome projects (1.2.3.1 Genome Projects in Humans), scientists have more profound understanding of human genome and the different sources of variation lying in it as well as their locations. The most frequent variations being SNPs, the third generation of genetic markers. SNP was first proposed to be used

for association mapping in 2000 [47] and the first SNP-based genome-wide association study (GWAS) appeared in 2005 where they conducted a genome-wide family-based association study based on a modified transmission disequilibrium test (TDT) using SNPs and found a haplotype associated with age-related macular degeneration [53].

However, the acknowledged starting point of GWAS is a publication by Wellcome Trust Case Control Consortium (WTCCC) in 2007 [54] because it is the first population-based association study using genome-wide marker data genotyped by high-coverage SNP chip [55]. In that study, 24 associations were reported for 7 diseases, including bipolar disorder, diabetes, coronary artery disease, etc.

Shortly after that, SNP-based GWAS have become the most popular mapping method all over the world and the increasing number of GWAS publications makes meta-GWAS possible. Meta-GWAS is a meta-analysis of the summary results from multiple GWAS for the same traits, which increases the detection power by merging the sample sizes with little computational cost [56]. The latest meta-GWAS for height by Genetic Investigation of ANthropometric Traits (GIANT) consortium has discovered ~700 height-associated loci using 253k genotyped individuals [57].

Thus far, ~55k SNP-trait-associations from more than 3k publications are catalogued in the online database, GWAS Catalog [58].

1.2.4 Relevant Researches and Future Researches

1.2.4.1 Biological Mechanism

Linkage and association studies localise the causal factors of diseases and traits on the genome (1.2.2.2 Linkage Mapping and 1.2.3.3 Association Mapping); however, they do not reveal the full detail of what these causal factors are, what these causal factors do and how these causal factors interact and result in disease. In fact, there is a huge gap from having the ‘bricks’ to building a ‘mansion’. For the public and clinical workers, the biological mechanism behind a complex trait or disease is much more important than the discovery of its causal factors as clinical prognosis, prevention, therapy and diagnosis largely rely on that knowledge.

To understanding the biological meaning of causal factors, one usually starts with gene annotation (gene ontology) [59]. By gene annotation, researchers could know the location of the QTLs in the genome and their nearby genes. Subsequently, it is possible to check the functions of these associated genes and whether it is plausible that their gene functions are related to the trait investigated (selecting the candidate genes). If a genetic variation happens to locate inside the protein-coding regions, it is possible to predict its function by checking whether it causes frame-shift, nonsense mutation, missense mutation, etc.

However, over 90% of the GWAS hits detected lie in non-coding regions. Studies on human traits and diseases show that mutations in non-coding regions could disrupt the binding of transcription factors and the functions of non-coding RNA can influence expression of active genes [60-62]. Therefore, further analysis is required for identifying the functions of causal variants, if they are non-coding variants.

One way is to consult the ENCyclopaedia Of DNA Elements database (ENCODE, https://genome.ucsc.edu/ENCODE/) to see whether the variants are within the regulatory elements of associated genes [63,64]. Another solution to understanding the role of non-coding variants is using genetically modified mice, i.e. modify or completely knockout the homologous gene in mice. The beauty of mouse study is we could observe the influence of a single mutation in any tissues at any growth stage, as long as the mutation is not lethal for mice and its effect is large enough. Mouse models of top GWAS hits for human diseases and traits is a trend in the post genomic era [65].

However, the real process of identifying causal variants is more complex because the identified genetic variant might not be the real causal variant. For example, the predicted function of an exonic variant might be wrong if the exonic variant is not the causal variant itself but is detected due to LD with that ungenotyped causal variant. Similarly, a variant found outside the gene could be tagging an ungenotyped variant within the gene. Briefly speaking, sequencing and re-sequencing techniques might help in such cases to identify potential causal variant for the identified signals but there are many other ways of exploring underlying causal variants in reality.

1.2.4.2 Application: Prediction, Prevention and Treatment

The purpose of studying a disease, certainly, is to optimise treatment. Clinical application is the most important thing that matters to the public. With the knowledge of trait causal factors and their mechanism learned from 1.2.2.2 Linkage Mapping, 1.2.3.3 Association Mapping and 1.2.4.1 Biological Mechanism, more potential options of therapeutic targets have been revealed [66,67]. Moreover, such knowledge is also useful for predicting disease risk and disease survival and selecting biomarkers for disease early diagnosis [68-72].

1.3 Lessons Learned from the History

1.3.1 My Understanding of Trait Architecture

Based on the review above, I think the architecture of a trait is the knowledge describing how trait compositions (causal factors) are bounded and formed into the outcome phenotype. Besides, the architecture of a trait is a time-specific and population-specific concept, i.e. trait architecture could change over generation and might differ across populations.

1.3.2 Five Important Types of Trait Architecture Study

Trait architecture studies can be classified into 5 categories. These are trait variance component analysis, detection of trait causal factors (or mapping studies), the influence of historical events and human behaviours on trait architecture, study of trait mechanism and clinical application. I call them the five key studies of trait architecture.

Here, I simply introduced the five key categories and briefly summarised the problems remaining in each of them. A more detailed methodology review for each type of trait architecture study is provided in the relevant chapters afterwards.

1.3.2.1 Trait Variance Component Analysis

Trait variance component analysis is a kind of study which explores the contribution of trait variance due to different origins, including additive genetics, dominance, epistasis, epigenetics, household effect, individual environment, etc. The appearance of trait variance component analysis could trace back to the beginning of 20th century (1.2.1.2 The Development of Statistical Genetics).

Heritability estimation study, the most frequent trait variance component analysis, could estimate the extent to which the trait variation is attributed by additive genetics. Thus, it provides us with an overall view of how heritable a trait or disease is and whether it is worthwhile to do genetic studies subsequently in the population.

It is possible to estimate the heritability of a trait by comparing the phenotypic resemblance between individuals with the genetic resemblance between individuals either inferred from the pedigree (pedigree based study such as MZ-DZ twin study [73,74]) or estimated using genotyped SNPs (such as genomic relationship matrix based restricted maximum likelihood, GREML [51,52]). However, the pedigree-based methods require relatives and thus the estimate of heritability might be biased due to confounding factors such dominance, epistasis, common environment, genetic-by-environment correlation and genetic-by- environment interaction shared among relatives if any [55,75,76]; whereas the GREML method is restricted by genotyping platform design, i.e. it is unable to capture causal genetic variants that are not tagged by SNP array, such as rare variants, copy number and structural variants [55,77,78].

A common problem for trait variance component analysis is that it usually focuses on estimating additive genetic variance but pays less attention to exploring other sources of influence such as dominance, epistasis and, more importantly, environmental factors. Although evidence shows that there probably is not much trait variance explained by genetic factors other than additive genetics [14], the summed effect of environment could still explain a very large proportion of trait variance. Taking different types of cancer as examples, the Nordic twin study demonstrates that the heritability estimates are low to moderate (9-58%) with mean of 33% across different cancers [79], which suggests that the majority of trait variance is due to environment.

1.3.2.2 Detection of Trait Causal Factors

Studies aimed at detecting trait causal factors, which are also known as mapping studies, reveal the location of the casual factors resulting in trait variation. Causal factor detection using molecular markers is a major force for trait architecture study which has started a long time ago.

A mapping study is guided by trait variance component analysis because the latter highlights the sources of trait causal factors and their contribution to the total trait variance. For example, GREML study shows that SNP-associated genetic variants explain ~50% trait variance for height [51], which suggests that: first, there are SNP-associated genetic variants for height; and second, if all SNP-associated genetic variants were detected, they should explain ~50% of the trait variance all together.

Currently, the most popular mapping method is GWAS, which has identified thousands of loci for a wide range of human complex traits and diseases [75,76]. However, the total heritabilities explained by identified hits are much lower than heritability estimates from pedigree-based or GREML studies due to platform design and the stringent significance tests applied [55,77,78].

There are more common problems for mapping studies. First, environment plays an important role in trait variance and gene-by-environment correlation might cause false discoveries in GWAS [80,81]. However, mapping studies for detecting environmental factors or mapping studies which considered the environmental influence are limited.

Second, irrespective of genetic mutation, a person's genome is fixed when the zygote is formed inside the mother's uterus and which set of the genes shall the zygote inherited from parents are almost randomly (genes might be co-segregated due to linkage) assigned during meiosis. Hence, genetic markers are perfect instrumental variables [82,83]. This means if a significant association between a trait and a genetic variant is found, then the cause-and-effect relationship is usually solid, i.e. the genetic variant (or an undetected genetic variant in strong LD with that detected genetic variant) must be the causation and its effect on that trait must be the causality. However, if the markers used for association studies are other omics data like gene expression levels and methylation rates or environmental records like smoking, then the cause-and-

effect relationship between trait and markers is undetermined. This also hinders the understanding of the mechanism for non-genetic causal factors.

Third, the property of omics data limits the findings. Using SNPs as genetic markers, it is only possible to discover causal genetic variants associated with SNPs; If researchers want to detect other types of genetic variants such as indel or structural variants (although some of them might be detected due to LD with SNPs), then they need to turn to other data such as sequencing [84]. If researchers want to explore the influence of chromatin interaction on a trait, they need to seek Hi-C data for help [85] or transcriptome data for splicing variation [86]. There are other omics data such as expression data, methylation data, proteomics data, etc. Different omics data serve different purposes and provide different information, it would be good to measure all of them but currently remains expensive. Furthermore, omics data will vary between tissue, over development and subject to environmental influences so measuring omics data in all potentially relevant circumstances (or even knowing what the relevant time and tissue are) might be impossible. For example, for a disease like cancer, somatic mutation and variation is important [87,88]. Therefore, to study the tumorigenesis and carcinogenesis of different cancers, researchers might want to sequence the DNA from different tissues.

1.3.2.3 Influence of Historical Events and Human Behaviours on Trait Architecture

The architecture of a trait is time specific and population specific as different historical events and human behaviours might change the trait architecture cumulatively over time and across populations.

Regarding genetic variants, they are maintained by mutation, fitness and mating. Mutation generates novel genetic variants, fitness decides whether a person could live long enough to produce offspring and mate choice determines what genes can be passed on to the next generation. Therefore, mutation, fitness and mate choice influence the gene pool and gene frequencies in the progeny and thus influence on them to some extent might alter trait architecture.

An example of mutation, the genetic mutations for the same trait might differ across populations because a mutation might raise in one population but not another. An example of fitness, the prevalence of sickle cell disease is much higher in African populations compared to non-African populations because the sickle gene has a protective effect against malaria (a common disease in Africa) [89]. Examples of mate choice, mating genetically related individuals (inbreeding) leads to inbreeding depression that increases the frequency of unfavourable homozygotes and, consequently, decreases the fitness of offspring [90]; whereas mating phenotypically similar individuals (assortative mating) [91,92] could increase the genetic variance of a trait by a maximum of twice [93].

Studies of the influence of historical events and human behaviours on trait architecture could trace back to the early 19th century (1.2.1.2 The Development of Statistical Genetics) and how they affect the genetic variance of a trait is in fact well-studied in population genetics [9,93,94]. However, population genetics focuses on the influence of human activities (input) on genetic variance and relatedness between relatives (outcomes); it does not reveal how these activities affect genetic variants (intermediates). With the popularisation of genotyping, we now have plenty of genetic information. How to integrate the influence of human behaviours and historical events into SNP-based models used for GWAS and GREML in order to explore the influence of human behaviours and historical events on genetic variants and trait architecture is a question that remains to be solved.

For the same trait, due to cultural, geographical, economic, historical and other reasons, different influential human activities might take place in the same population at different times and in different populations at the same time. Therefore, it is important to know the difference in historical events and human behaviours between two different populations (or two different generations) and use such information to interpret the similarities and difference in trait architecture across populations (or across time) because historical events and human behaviours are the direct causation of heterogeneity. However, owing to how data are usually collected (i.e. at a certain time point and within a population), such cross-population or cross-generation comparison study is seldom seen.

1.3.2.4 Study on Functional Pathways and Networks

DNAs are transcribed into RNAs, mRNAs are translated into proteins, proteins and other biomolecules interact and form a biological pathway and several biological pathways interact and result in the final phenotype we observed, along with other factors such as the regulation of gene expression, the influence of environmental factors on the pathways, etc. This probably is the simplest interpretation explaining how trait causal factors are connected and forming the biological pathways and networks underlying trait variation.

Currently, the study of trait mechanisms emphasises on both ends of the biological mechanism network. From the DNA side: discovering how genetic variants detected in a mapping study affect the trait and how they are regulated by other genetic and epigenetic variants. And from the phenotype side: such as researching the influence of alcohol metabolism on alcohol intoxication [95]. However, the link between two ends is usually missing.

To weave the complete network of the biological mechanism of a complex trait or disease, we need multi-disciplinary cooperation, which might include quantative genetics, molecular genetics, epigenetics, cellular biology, metabolomics, embryology, physiology, ecologist, etc. Therefore, emphasis should be paid on integrating the results learned from different biological disciplines for the same trait together, trying to draw a full picture of how trait casual factors interact and forming the final phenotype. And a better understanding of trait mechanism network, in return, would help to detect those causal factors which are missed in mapping studies but shown up in the mechanism network [96,97].

1.3.2.5 Clinical Application

Clinical application is the final step of a trait architecture study and the outcome that matters most to the public. However, the conversion rate from scientific research into clinical application is low.

For example, GWAS marks the location of trait-associated loci in the genome, but the region represented by one QTL could be wide enough to harbour multiple genes; Or

no genes at all as most GWAS hits lying in non-coding regions [60]. Therefore, it is not obvious which is the functional variant contributed to the GWAS signal [98]. Consequently, only a very small fraction of GWAS hits are found in the current drug target database [66]. To discover novel drug targets, we still require the knowledge of gene regulatory network (GRN) as studies show that genes closely correlated with GRN of published GWAS hits look promising for drug targets [66,67].

Regarding prediction, the overall prediction accuracy for human complex traits is low, e.g. the prediction accuracies for height and BMI using genotype and methylation data in unrelated individuals are no more than 20% [99]. Since environmental factors are another major component of trait variance and theoretically prediction accuracy should benefit from adding environmental causal factors, future prediction studies might consider taking them into account.

1.4 Aims and Thesis Outline

1.4.1 Aims

The aims of this project are two.

Aim one, exploring trait architecture for human complex traits, going through the five key trait architecture studies one by one, including trait variance component analysis, trait causal factor detection, influence of historical events and human behaviours on trait architecture, biological pathways and clinical application. The complex traits investigated here are anthropometric and cardio-metabolic traits such as height, body mass index (BMI) and high-density lipoprotein (HDL), detailed review for those traits is provided in Chapter 2.

Aim two, while conducting analyses for each type of trait architecture study, trying to solve the current issues remaining in each field or improving the current methods frequently used. Therefore, the studies in my thesis are not simply copy and paste, i.e. applying the methods used in published studies into our data “as is”, but also attempting to develop novel analytical approaches to dissecting out the causes of variation between individuals.

1.4.2 Outline of the Thesis

During my 4-year PhD study, I have conducted four out of the five key trait architecture studies mentioned above, which are trait variance component analysis, trait causal factor detection, studying the influence of historical events and human behaviours on trait architecture and clinical application.

In Chapter 2, I conducted variance component analysis. In contrast to traditional heritability, I also explored the influence of familial environmental factors, such as nuclear family environment, couple environment and sibling environment on trait variance explanation.

In Chapter 3 and 4, I conducted a GWAS and a prediction study respectively. In contrast to traditional GWAS and prediction studies, my GWAS and prediction models include the important non-genetic sources of trait variance found in Chapter 2

In Chapter 5, I explored the influence of assortative mating on trait architecture. I developed a novel pedigree study to estimate assortative mating intensity (the strength of phenotypic assortation) and heritability, using the expected resemblances between relatives and between in-laws.

At the end, I summarised and discussed my project in Chapter 6.

Chapter 2: Trait Variance Component Analysis

2.1 Introduction

In Chapter 1, I brought out the five key studies in the exploration of trait architecture, the first among which is trait variance component analysis. Trait variance component analysis reveals the proportion of trait variance contributed by different sources of variation, which precedes the follow-up trait architecture studies. One of the issues remaining in trait variance component studies, as mentioned in Chapter 1, is ignoring environmental contributions. Therefore, an important aspect of this research is to model the contributions of environmental effects shared within family on trait variance and tease apart those environmental effects from the confounding genetics associated with the pedigree to enable assessment of the latter, i.e. how much genetic variance is not captured by SNPs.

Currently, large-cohort studies are a common trend in the genetic field. Inevitably, it is possible to recruit participants with known or unknown multiple degrees of relationship in the study. Therefore, confounding factors shared among relatives such as household effects might be present in the data structure. Removing confounding factors by means of taking related individuals completely out of the study greatly reduces the sample size, as well as the chance to fully investigate the causes of variation in the whole population, and thus is inadvisable. To accommodate large-cohort studies with multiple degrees of relatives and make full use of the data, new methods are required.

Consequently, in this chapter, I developed novel trait variance component analyses which model confounding environmental and genetic factors in the presence of relatives. In my model, trait variation is dissected into SNP-associated genetic effects, pedigree-associated genetic effects, nuclear family environment, couple environment, sibling environment and residuals. SNP-associated genetic effects refer to genetic variation attributed by common variants inherited from distant ancestors that are in linkage disequilibrium (LD) with markers on the SNP array at the population level;

Pedigree-associated genetic effects refer to genetic variation contributed by rare variants, copy number variants (CNVs), structural variants, etc. that cluster in specific families and are captured due to strong linkage in high-order pedigrees but are not in population-wide LD with common SNPs; nuclear family environment refers to environmental variation due to common environmental factors shared among nuclear family members (parents-offspring, siblings and couples) such as the household effect; couple environment refers to common environmental factors shared by partners (member of a couple) such as diet; common environmental factors shared between full-siblings such as rearing environment; and any other genetic or environmental factors unaccounted (the residual) in the model respectively. These five effects are the key elements in this thesis as they will be mentioned repeatedly in this and other chapters.

My method was validated by simulation study. Subsequently, I applied my approach to 16 traits related to anthropometrics and cardio-metabolism, such as height, BMI and HDL, in a cohort consisting of ~20k individuals with recent Scottish descent. Details about methodology, simulation study, cohort description and results, as well as a review of trait component analysis and traits investigated, are published in a paper titled "Pedigree- and SNP-Associated Genetics and Recent Environment are the Major Contributors to Anthropometric and Cardiometabolic Trait Variation" in PLOS Genetics [100]. Here, I attached the manuscript as it was accepted by the journal in this chapter and attached the supplementary materials in Appendix. However, to comply with the naming of figures and tables through the whole thesis and thesis format, the title indexes, font, etc., were changed.

2.2 Publication: Pedigree- and SNP-Associated Genetics and Recent Environment are the Major Contributors to Anthropometric and Cardiometabolic Trait Variation

2.2.1 Abstract

Genome-wide association studies have successfully identified thousands of loci for a range of human complex traits and diseases. The proportion of phenotypic variance explained by significant associations is, however, limited. Given the same dense SNP panels, mixed model analyses capture a greater proportion of phenotypic variance than single SNP analyses but the total is generally still less than the genetic variance estimated from pedigree studies. Combining information from pedigree relationships and SNPs, we examined 16 complex anthropometric and cardiometabolic traits in a Scottish family-based cohort comprising up to 20,000 individuals genotyped for ~520,000 common autosomal SNPs. The inclusion of related individuals provides the opportunity to also estimate the genetic variance associated with pedigree as well as the effects of common family environment. Trait variation was partitioned into SNP-associated and pedigree-associated genetic variation, shared nuclear family environment, shared couple (partner) environment and shared full-sibling environment. Results demonstrate that trait heritabilities vary widely but, on average across traits, SNP-associated and pedigree-associated genetic effects each explain around half the genetic variance. For most traits the recently-shared environment of couples is also significant, accounting for ~11% of the phenotypic variance on average. On the other hand, the environment shared largely in the past by members of a nuclear family or by full-siblings, has a more limited impact. Our findings point to appropriate models to use in future studies as pedigree-associated genetic effects and couple environmental effects have seldom been taken into account in genotype-based analyses. Appropriate description of the trait variation could help understand causes of intra-individual variation and in the detection of contributing loci and environmental factors.

2.2.2 Keywords

GREML; Anthropometric Traits; Cardiometabolic Traits; Heritability; Common SNPs; Pedigree Study; Variance Component Analysis; Family Environment; Couple Environment; Sib Environment; Height; Body Mass Index; Weight; Waist; Hips; Waist-Hips Ratio; A Body Shape Index; Fat; Glucose; Creatinine; High Density Lipoprotein; Urea; Total Cholesterol; Systolic Blood Pressure; Diastolic Blood Pressure; Heart Rate;

2.2.3 Author Summary

Unravelling overall trait architecture of complex traits and diseases is important for phenotype prediction and disease prevention and correct modelling of the trait will further aid discovery of causative loci. Here we take advantage of genome-wide data and a large family-based study to examine the role of common genetic variants, pedigree-associated genetic variants, shared family environment, shared couple environment and shared sibling environment on 16 anthropometric and cardiometabolic traits. By analysing up to ~20,000 Scottish individuals, we find that common genetic variants, pedigree-associated genetic variants and recently-shared environment of couples are the most important contributors to variation in these traits, while past family and sibling environment have a limited impact. Further studies on the pedigree-associated genetic variation and the shared couple environment effect are needed, as little research has been devoted to them so far.

2.2.4 Introduction

Phenotypic variation for a quantitative trait is attributable to the summed effects of genetic and environmental influences together with any covariances and interactions. The proportion of phenotypic variance contributed by genetic variation is termed the heritability (h^2) [93]. The heritability scales the influence of genetic and environmental factors on phenotypic variation. This provides us with insights into the genetic and environmental architecture of human complex traits and our potential

ability to dissect out loci associated with trait variation and is also useful for the prediction of heritable disease risk [80,81]. As a consequence, such knowledge is of potential value for clinical diagnosis, therapy, prevention and prognosis [101]. Therefore, obtaining unbiased estimates of variation caused by different factors and the heritability of traits relevant to health and disease processes is important.

A classic approach to gauging the heritability in humans is by comparing the observed phenotypic similarity to the expected genetic resemblance between relatives inferred from family pedigrees [102]. This method evaluates the pedigree based heritability (h^2_{ped}) indirectly without requiring information on the inheritance of individual loci and thus, is quite practical and still widely-used in twin, family and other pedigree studies [73,74]. Note that, h^2_{ped} is often considered to be an estimate of the true heritability h^2. Genome-wide association studies (GWAS), on the contrary, identify causal loci through their association with recorded genetic markers and then aggregate the proportion of variance explained by statistically-significant variants [103,104], which is sometimes referred to as the "GWAS heritability" (h^2_{GWAS}). Each approach has its limitations and drawbacks. Pedigree studies require genealogical information from known relatives to deduce their expected genetic resemblance and h^2_{ped} may be biased due to the factors shared among relatives (including dominance, epistasis, common environment, genetic-by-environment correlation and genetic-by-environment interaction) if such effects are present and the available pedigree structure does not allow these to be accounted for in the analysis [55,75,76]. Although GWAS have been very successful at discovering novel loci for a range of polygenic disease and complex traits, they have been less successful at capturing the full extent of known trait genetic variance [75,76]. This is probably because of their failure to detect particular types of variants such as common variants with small effects, rare variants, copy number variants and structural variants, as a consequence of inadequate sample size, genotyping platform design and analyses used, together with the stringent statistical tests applied [55,77,78]. As a result, there usually is a substantial gap between the estimates of h^2_{ped} and h^2_{GWAS}, often termed the "missing heritability" [75,105].

Recently, Yang et al. [51,52] have championed an approach, known as GREML [106], to estimate the amount of trait variance explained by SNPs. The estimation of the SNP (or genomic) heritability (h_g^2), which refers to the additive genetic effects captured by genotyped SNPs, utilises a matrix comprising realised genetic relationships inferred from genomic marker data originally gathered for GWAS (known as genomic relationship matrix or GRM) [51,52]. The h_g^2 estimate from this approach, when estimated using unrelated individuals, lies between the h_{ped}^2 and h_{GWAS}^2 estimates, and has been considered as a lower limit for the former and an upper limit for the latter [75,76]. As an example, for height, h_{GWAS}^2, h_g^2 and h_{ped}^2 from three different studies are 0.10, 0.45 and 0.80 respectively [51,102,103]. This suggests that a substantial proportion of the genetic contribution to trait variation is SNP-associated and hence contributes to h_g^2 but not all this variation is detected by current GWAS, probably due to a combination of insufficient sample size and stringent significant thresholds employed. The difference between h_g^2 and h_{ped}^2 may be largely due to trait associated variants not in linkage disequilibrium (LD) with genotyped SNPs, such as rare variants, copy number variants (CNV) and other structural variants as mentioned above. Variation associated with such effects is captured by h_{ped}^2 due to strong LD in relatives [107].

Recent studies have started dissecting the heritable component of variation and other components shared among relatives by studying more complex populations made-up of both unrelated individuals and extended pedigrees [75,76,107]. For instance, Zaitlen et al. [76] have demonstrated that simultaneously including in a GREML analysis a GRM and a modified GRM (in which entries smaller than a certain threshold in the GRM are set to zero) can be used to jointly estimate SNP-associated and total heritabilities in the presence of relatives. We also note that shared environment may be an important contributor to heritability inflation when close relatives are included in analysis.

In this study, we use data from a single homogeneous cohort consisting of approximately 20,000 adults with varying degrees of relationships sampled from Scotland. The individuals have data on over 520,000 SNPs distributed across the autosomes. The dense marker information together with extended genealogical

information allows us to partition the phenotypic variance and explore the genetic and environmental effects shared among related individuals (both biological relatives and couples).

We analyse eight anthropometric traits, comprising height, weight, fat, body mass index (BMI), hips, waist, waist-to-hips ratio (WHR) and a body shape index (ABSI) [108] and eight cardiometabolic traits, comprising levels of creatinine, urea, total cholesterol (TC) and high density lipoprotein (HDL) in serum, level of glucose in blood, systolic blood pressure (SBP), diastolic blood pressure (DBP) and heart rate (HR).

In our work, we implement alternative models to estimate effects that might contribute to the variation in the 16 traits analysed. Results show that, with these data, we can separate total genetic variation into SNP-associated and pedigree-associated genetic influences. We also observe that past family environment and shared full-sibling environment generally have a limited impact on trait variation, whereas the effect in couples of living in the current (shared) environment is always important in our data.

2.2.5 Results

2.2.5.1 Overview of the Methods

We conducted variance component analyses to dissect the phenotypic variation for traits recorded in the Generation Scotland: Scottish Family Health Study (GS:SFHS) cohort [21] into genetic and environmental factors. Analyses utilised a mixed-model approach implemented in a restricted maximum likelihood (REML) framework using the GCTA software [52]. The population was divided into two tranches of approximately equal size and genotyped in two stages. All initial analyses were performed with the first 10,000 genotyped individuals, (named GS10K). GS10K comprised small nuclear families (largely two parents and two offspring) together with unrelated individuals, although inevitably there were second degree and more distant relationships included. The second tranche completed the genotyping of the rest of the population (another 10,000 individuals) including further relatives in incomplete families (e.g. missing samples from parents and additional siblings, as well as other

relationships), resulting particularly in a proportional increase in the number of second and third degree relationships (Table 2.1). To confirm results obtained from GS10K, some of the analyses were repeated in the whole 20,000 individual sample (named GS20K).

Table 2.1 Comparisons of sample sizes, number of non-zero off-diagonal entries and number of pairwise relationships of different degrees between GS10K and GS20K.

	GS10K	**GS20K**	**Ratio**
No. IDs	9,863	20,032	1 : 2.03
Matrix	**No. Non-Zero Off-diagonal Entries** [a]		**Ratio**
G	48,634,453 [b]	200,630,496	1 : 4.13
K	8,080	41,174	1 : 5.10
F	4,821	20,115	1 : 4.17
C	1,283	1,767	1 : 1.38
S	676	8,495	1 : 12.57
Degree of Relationship [c]	**No. Pairs**		**Ratio**
1st degree relatives	3,529	18,320	1 : 5.19
2nd degree relatives	441	7,851	1 : 17.80
3rd degree relatives	500	4,129	1 : 8.26
4th degree relatives	1,099	3,950	1 : 3.59
5th degree relatives	3,891	11,032	1 : 2.84
unrelated individuals	48,624,993	200,585,162	1 : 4.13

[a] The number of off-diagonal entries is calculated in the lower triangular part of all the matrices

[b] For matrix G, all the off-diagonal entries are different from zero, so the value represents the total number of off-diagonal entries

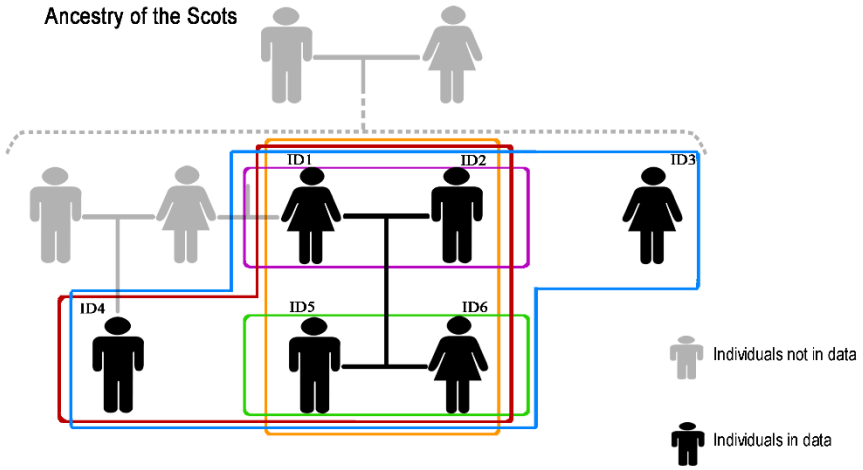
[c] Distance of relationship is identified according to an approximate range of the expected pair-wise relatedness, $0.5^{i-0.5}$ to $0.5^{i+0.5}$ for i^{th} degree relatives.

We first explored the extent to which estimates of h_g^2 were inflated by the inclusion of relatives. We subsequently analysed our data allowing trait variation to be potentially influenced by both genetic and environmental effects. We assumed that the genetic effects comprised additive genetic effects associated with genotyped SNPs (h_g^2) and additional additive genetic effects associated with pedigree but not with genotyped SNPs (h_{kin}^2), and we assumed that the environmental effects potentially comprised nuclear family effects (e_f^2) common to both parents and offspring, full-sibling effects (e_s^2) common to just siblings and couple effects (e_c^2) common to just the members of a couple (Figure 2.1). The total heritability, termed h_{gkin}^2 in this manuscript, referred to as $h_{IBS>t_*}^2$ in Zaitlen et al. [76] and comparable to h_{ped}^2 from traditional pedigree studies, was estimated as the sum of h_g^2 and h_{kin}^2 for each model. To allow estimation of the influence of each effect, we generated five design matrices: $\mathbf{GRM_g}$, $\mathbf{GRM_{kin}}$, $\mathbf{ERM_{Family}}$, $\mathbf{ERM_{Sib}}$ and $\mathbf{ERM_{Couple}}$ respectively, where GRM refers to genomic relationship matrices and ERM refers to environmental relationship matrices.

For brevity, we named different alternative models using abbreviations according to first subscript letter of the effects examined. We coded 'G' for $\mathbf{GRM_g}$, 'K' for $\mathbf{GRM_{kin}}$, 'F' for $\mathbf{ERM_{Family}}$, 'S' for $\mathbf{ERM_{Sib}}$ and 'C' for $\mathbf{ERM_{Couple}}$ –e.g. model 'GKC' = $\mathbf{GRM_g + GRM_{kin} + ERM_{Couple}}$. All models included a residual matrix (allowing effects specific to an individual that were not shared with any other member of the population).

We identified the most appropriate model for each trait by a stepwise model selection process via removing non-significant components from the full model based on a Wald test of their estimated effect and a likelihood ratio test (LRT), and we estimated the effects of significant factors using the selected models in GS10K. We repeated the model selection and corresponding variance component analyses in GS20K to identify differences resulting from analysing a more complex population structure, encompassing a larger proportion of close relationships.

Figure 2.1 Illustration of the model and matrices



h^2_g: The total additive genetic effects captured by genotyped SNPs. (Variance: σ^2_g ; Matrix: GRM_g)

	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6
ID 1	1.001	-0.012	0.003	0.220	0.496	0.499
ID 2	-0.012	0.998	0.011	0.007	0.505	0.502
ID 3	0.003	0.011	1.015	0.009	0.001	0.002
ID 4	0.220	0.007	0.009	1.003	0.115	0.136
ID 5	0.496	0.505	0.001	0.115	0.979	0.578
ID 6	0.499	0.502	0.002	0.136	0.578	1.007

e^2_f: The common environmental effects shared by family members of a nuclear family. (Variance: σ^2_{ef} ; Matrix: ERM_{Family})

	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6
ID 1	1	1	0	0	1	1
ID 2	1	1	0	0	1	1
ID 3	0	0	1	0	0	0
ID 4	0	0	0	1	0	0
ID 5	1	1	0	0	1	1
ID 6	1	1	0	0	1	1

h^2_{kin}: The extra genetic effects associated with pedigree. (Variance: σ^2_{kin} ; Matrix: GRM_{kin})

	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6
ID 1	1.001	0	0	0.220	0.496	0.499
ID 2	0	0.998	0	0	0.505	0.502
ID 3	0	0	1.015	0	0	0
ID 4	0.220	0	0	1.003	0.115	0.136
ID 5	0.496	0.505	0	0.115	0.979	0.578
ID 6	0.499	0.502	0	0.136	0.578	1.007

e^2_c: The common environmental effects shared by a couple..(Variance: σ^2_{ec} ; Matrix: ERM_{Couple})

	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6
ID 1	1	1	0	0	0	0
ID 2	1	1	0	0	0	0
ID 3	0	0	1	0	0	0
ID 4	0	0	0	1	0	0
ID 5	0	0	0	0	1	0
ID 6	0	0	0	0	0	1

Residuals: The remaining genetic and environmental effects after accounting for h^2_g, h^2_{kin}, e^2_f, e^2_c and e^2_s. (Variance: σ^2_e ; Matrix: I)

	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6
ID 1	1	0	0	0	0	0
ID 2	0	1	0	0	0	0
ID 3	0	0	1	0	0	0
ID 4	0	0	0	1	0	0
ID 5	0	0	0	0	1	0
ID 6	0	0	0	0	0	1

e^2_s: The common environmental effects shared by full-siblings, which may include non-additive genetics. (Variance: σ^2_{es} ; Matrix: ERM_{Sib})

	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6
ID 1	1	0	0	0	0	0
ID 2	0	1	0	0	0	0
ID 3	0	0	1	0	0	0
ID 4	0	0	0	1	0	0
ID 5	0	0	0	0	1	1
ID 6	0	0	0	0	1	1

The diagram shows the relationship between the tested genetic/environmental effect and the individuals in an example pedigree. Each colour represents a specific effect and individuals affected by that effect are circled with that colour. People in grey or black are the people not in or in the data. Examples of how the relationship matrices for those effects look are also given.

More details about traits, matrices and models are given in Material and Methods and Table S2.1 and Table S2.2. In the main manuscript, we only list results for the final models identified by the model selection procedure and the full model, but a comprehensive list of estimates obtained for the different effects for each trait and each model is available in Table S2.3 and Table S2.4.

Model robustness and the effectiveness of the model selection were tested using simulated data based on GS10K.

2.2.5.2 Simulation Study: Robustness of the Models

We conducted a simulation study using real genotype and pedigree information from GS10K to evaluate the robustness of our models. To make computation feasible, we mainly focused on data simulated under the simplest and most complex models (models 'G', 'K', 'F', 'S', 'C', 'GK', 'GF' and 'GKSC') and those representing the commonest conclusions of model selection in analyses of the real GS10K data (models 'GF', 'GFS', 'GKC' and 'GKSC'). Table S2.5 shows the simulated and observed values for each parameter as well as the model we used for analyses in different scenarios.

In the first scenario, we examined the performance of our models (models 'G', 'K', 'F', 'S' and 'C') when simulated phenotypes were only contributed by one of the five corresponding effects plus residual variation. Under these models (Table S2.5), the mean of overall estimates per parameter was very close to its simulated value, indicating that our design matrices $\mathbf{GRM_g}$, $\mathbf{GRM_{kin}}$, $\mathbf{ERM_{Family}}$, $\mathbf{ERM_{Couple}}$ and $\mathbf{ERM_{Sib}}$ worked well in simple models and were able to capture their corresponding effects even when the simulated variance associated with an effect was low (≤ 3%).

In the second scenario, we evaluated the performance of our models (models 'GK' and 'GF') when the simulated phenotypes were determined by SNP-associated genetic effects and one of the familial effects (either pedigree-associated genetics or nuclear family environment) plus residual variation. Results (Table S2.5) indicate that, in cohort with familial structure, failure to account for or inaccurate modelling of familial effects (i.e. when models used were inconsistent with phenotypic contributors) would

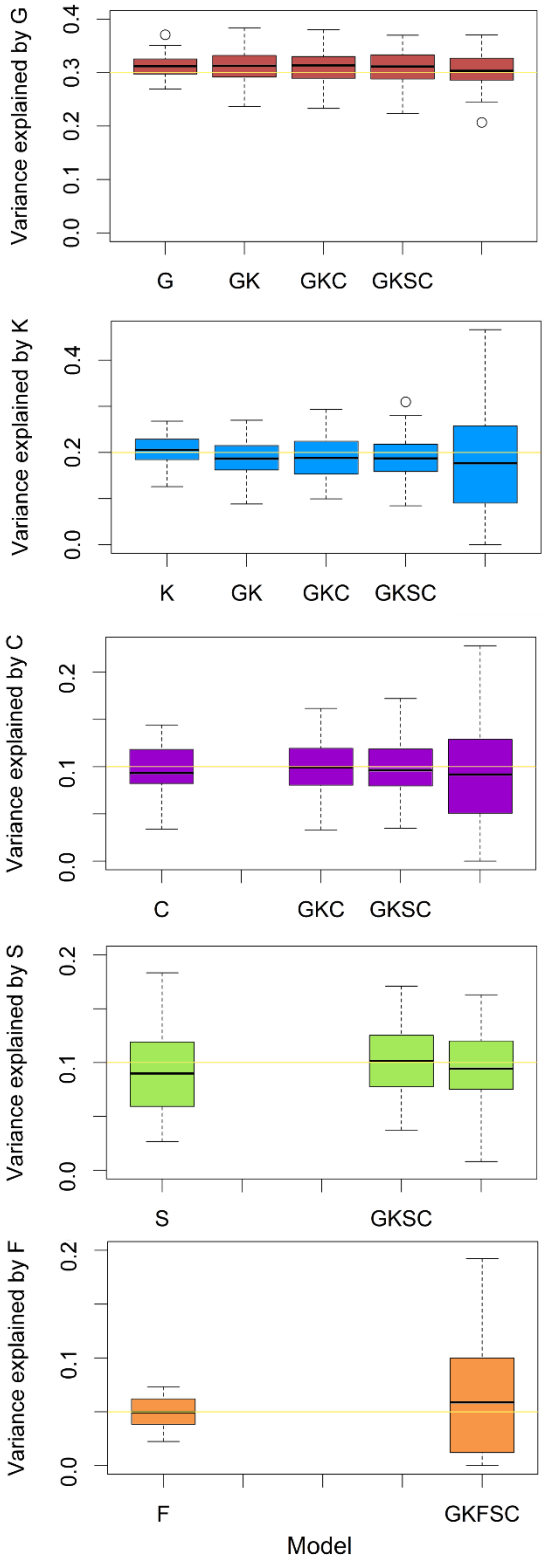
result in upward bias for h_g^2 in the presence of relatives. However, this upward bias due to the confounding familial factors could be eliminated by either excluding nominally related individuals or using the appropriate models for analysis. The former method removes the ability to estimate the familial effects as well as reducing the sample size, whereas using the appropriate models, estimates obtained were very close to their parameter settings and gave a good idea of the magnitude and approximate values of SNP and familial effects as well as the total proportion of variance explained by additive genetics ($h_{gkin}^2 = h_g^2 + h_{kin}^2$), despite the fact that the means of estimates of h_g^2, h_{kin}^2 and e_f^2 were usually significantly different from the original parameter settings.

In the third scenario, we inspected the performance of the full model 'GKFSC' and models selected from analyses of real phenotypes in GS10K other than 'GF' (models 'GFS', 'GKC' and 'GKCS'). Results (Table S2.5) demonstrate that all models were robust in terms of the mean of overall estimates per parameter being either unbiased or very close to original settings.

Figure 2.2 summarizes the main results from these simulations, showing the overall performance of our design matrices from simple models to complex models.

The median of estimates for each component was unbiased across simple and complex models, however, the estimates for h_{kin}^2, e_f^2 and e_c^2 were quite variable in the full model, probably due to limitations imposed by the data structure. All of the above verify the robustness of our models.

Figure 2.2 Boxplots for estimates of each component obtained from models 'G', 'K', 'F', 'S', 'C', 'GK', 'GKC', 'GKSC 'and 'GKFSC'



X-axis: the contributors to the simulated phenotype and the model used (matched model); Y-axis: proportion of total phenotypic variance captured by each design matrix. Yellow lines: simulated value for each component. Parameter settings: h_g^2 =0.3, h_{kin}^2 =0.2, e_c^2 =0.1, e_s^2 =0.1 and e_f^2 =0.05. For example, the 2nd boxplot of the 3rd graph means that, the simulated phenotypes are contributed by 30%, 20%, 10% and 40% of SNP-associated, pedigree-associated, couple environmental and residual effects respectively; we conducted variance component analyses for all replicates using the matched model **'GKC'** and the estimates of e_c^2 range from about 8% to 12% with a mean of 10%, as expected.

2.2.5.3 Simulation Study: Effectiveness of the Model Selection Procedure

Although we confirmed that our models were robust (Table S2.5 and Figure 2.2), the potentially high correlation between $\mathbf{ERM_{Family}}$ matrix and combined $\mathbf{ERM_{Couple}}$ and $\mathbf{GRM_{kin}}$ matrices may make it challenging to jointly estimate h^2_{kin}, e^2_f and e^2_c accurately in our sample as the standard errors for those parameter estimates obtained from the full model were high (Table S2.4). Thus the most challenging part of our study may be to precisely dissect pedigree-associated genetic effects, shared nuclear family environment and shared couple environment. Therefore, we performed model selection using simulated data to test our model selection procedure where simulated phenotypes were contributed by moderate SNP-associated genetic effects and low sibling environmental effects plus a) moderate nuclear family environmental effects but low pedigree-associated genetic effects and couple environmental effects; b) low nuclear family environmental effects but moderate pedigree-associated genetic effects and couple environmental effects; or c) moderate nuclear family environmental effects, pedigree-associated genetic effects and couple environmental effects. All scenarios included residual variation.

Table S2.6 shows the parameter settings and the summary of model selection procedure performance for these scenarios. We expected that our model selection procedure was able to identify SNP genetics ($\mathbf{GRM_g}$) and nuclear family environment ($\mathbf{ERM_{Family}}$) or SNP and pedigree genetics ($\mathbf{GRM_{kin}}$) and couple environment ($\mathbf{ERM_{Couple}}$) or SNP and pedigree genetics and nuclear family and couple environment accordingly, since they were the major factors in each corresponding scenario.

As results demonstrated, in all situations our model selection procedure generally ($\geqslant$ 80%) selected the appropriate model which contains all major components of phenotypic variation. The remaining times in the first two of these scenarios, pedigree-associated genetic effects or those plus shared couple environment were selected instead of nuclear family environmental effects or vice versa, and in the remaining two replicates in the third of these scenarios we missed pedigree-associated genetic effects. In addition, our model selection never fully detected all minor contributions to the phenotype in the first two of these scenarios when the minor effects were too small (e.g. effects contribute to ≤5% of the phenotypic variance).

Both issues identified above (~20% chance of selecting inappropriate models and failure to identify all minor effects) are likely to have been due to limitations in the data structure of GS10K, which provides too few of the appropriate relationships for corresponding effects (pedigree-associated genetics, nuclear family, sibling and couple environment) to resolve correlations between parameters and detect minor effects. These limitations have been greatly ameliorated in the GS20K data.

We also conducted variance component analyses using the final selected model for each replicate (Table S2.6). For those replicates that had appropriate models after model selection, the estimates of factors that remained in the models were usually close to, and not significantly different from, their simulated values, indicating that the results from selected models were reliable. More details about simulation study can be found in Text S2.1, Table S2.5 and Table S2.6.

2.2.5.4 Impact of Inclusion of 1st Degree Relatives on the Genomic Heritability in GS10K

In the first analyses of the real data, we looked for evidence of familial effects (either pedigree-associated genetics or nuclear family environment) in our cohort. As shown by simulation (Table S2.5), if there were any familial effects, we should obtain inflated estimates of h_g^2 when we conducted variance component analyses using model 'G' in the presence of relatives, compared to the estimates of h_g^2 given from the unrelated subpopulation. GS10K consists of nearly 10,000 genotyped individuals with multiple degrees of relationship, which allows us to explore the impact of familial effects on h_g^2 estimation in this cohort.

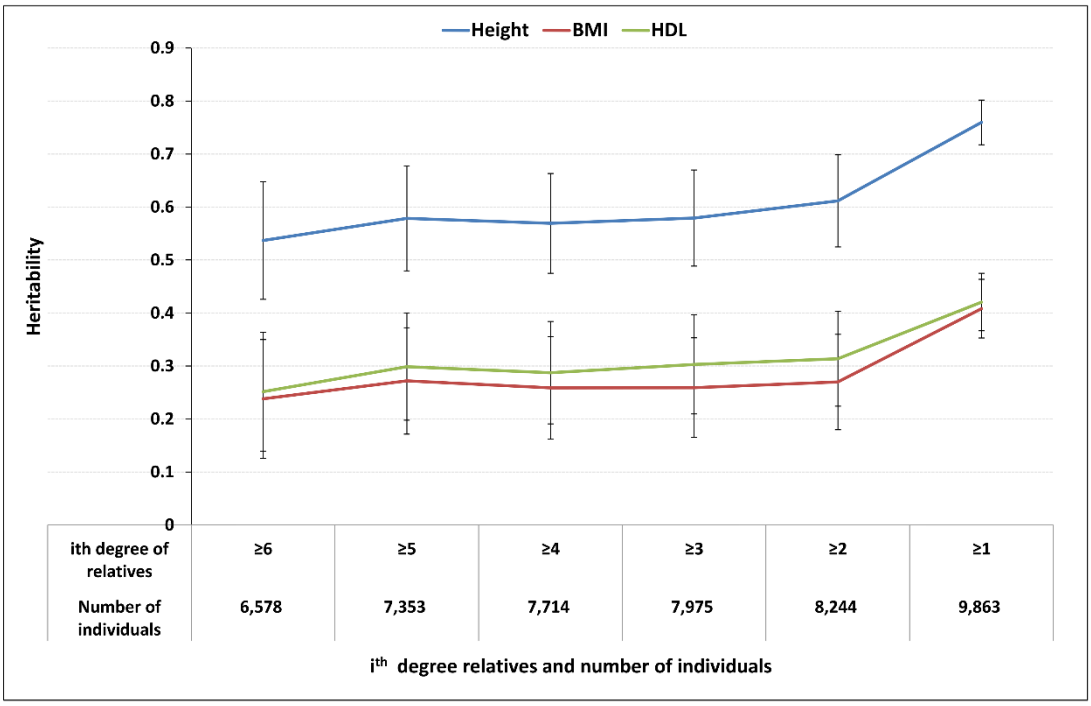
Table 2.1 shows the population structure of genotyped individuals in GS10K. The degree of relationship between two individuals was identified according to an approximate range of the expected pair-wise relatedness (r), which was from $0.5^{i-0.5}$ to $0.5^{i+0.5}$ for i^{th} degree relatives (e.g. pairs of individuals with relatedness from 0.354 to 0.707 were considered as 1st degree relatives).

With these criteria, GS10K consisted of more than 3,500 pairs of 1st degree relatives, around 450 pairs of 2nd and 500 pairs of 3rd degree relatives, but the majority of pairs

of individuals (over 99.9%) were genetically unrelated (more distant than 5th degree relatives, $r \leq 0.022$). In total, there were around 6,600 unrelated individuals (defined using the criteria described above) in GS10K.

We estimated h_g^2 for each trait using model 'G' for subpopulations of GS10K made-up of individuals with different degrees of relatedness (using the upper bound of the expected relatedness of each category as GRM cut-off points in GCTA). Figure 2.3 shows how h_g^2 estimates for height, BMI and HDL changed as we progressively included more closely related individuals in the relationship matrix. Results for the remaining traits are shown in Table S2.3.

Figure 2.3 Heritability estimates using subpopulations of GS10K with different GRM cut-off points.



X-axis: the number of individuals among whom the pairwise relationship is larger than a specified degree; Y-axis: Heritability estimates with ± 2 S.E.

In general, h_g^2 estimates were stable as we gradually added more closely related individuals in the analyses until the inclusion of 1st degree relatives that resulted in inflation of the estimates (Figure 2.3 and Table S2.3). Based on our results, h_g^2 was overestimated only when 1st degree relatives were included. For glucose and DBP, the h_g^2 estimates did not appear inflated after 1st degree relatives were included, suggesting that these traits were not affected by familial effects (Table S2.3).

2.2.5.5 Variance Component Analyses using the Full Model 'GKFSC' and Stepwise Model Selection in GS10K

The increase in h_g^2 estimates resulting from the inclusion of 1st degree relatives provided evidence of familial variation in our cohort. However, it is not clear whether these familial effects are due to pedigree-associated genetic effects or shared nuclear family environment or both because either of them has the ability to inflate h_g^2 estimates (this was also observed in the simulation data: Table S2.5: scenario ii). Therefore, we attempted to tease out this familial variance from the total phenotypic variance and dissect the familial variation as well as the remaining trait variation further using the full model 'GKFSC' and the stepwise selection procedure to define a final model containing the most important effects contributing to trait variation.

Table 2.2 shows the results for final models selected from stepwise model selection strategies and for the proportions of total phenotypic variance explained by different effects using final models, as well as for those obtained using the full model.

Table 2.2 Results of variance component analyses for anthropometric and cardiometabolic traits using final models selected from the stepwise model selection and the full model in GS10K

Trait	Model		GRM_g h^2_g (s.e.)	GRM_{kin} h^2_{kin} (s.e.)	ERM_{Family} e_f^2 (s.e.)	ERM_{Sib} e_s^2 (s.e.)	ERM_{Couple} e_c^2 (s.e.)
Anthropometric Traits							
Height	Selected	**GKC**	0.47(0.04)	0.36(0.05)			0.16(0.03)
	Full	**GKFSC**	0.45(0.04)	0.36(0.17) NS	0.00(0.08) NS	0.02(0.03) NS	0.17(0.09) NS
Weight	Selected	**GF**	0.28(0.03)		0.18(0.02)		
	Full	**GKFSC**	0.28(0.04)	0.17(0.17) NS	0.10(0.09) NS	0.01(0.04) NS	0.09(0.09) NS
Fat	Selected	**GKC**	0.26(0.04)	0.26(0.06)			0.19(0.03)
	Full	**GKFSC**	0.26(0.04)	0.21(0.18) NS	0.02(0.09) NS	0.02(0.04) NS	0.16(0.09)
BMI	Selected	**GKC**	0.25(0.04)	0.33(0.05)			0.21(0.03)
	Full	**GKFSC**	0.25(0.04)	0.19(0.17) NS	0.07(0.09) NS	0.00(0.04) NS	0.14(0.09)
Hips	Selected	**GKC**	0.21(0.04)	0.27(0.06)			0.17(0.03)
	Full	**GKFSC**	0.21(0.04)	0.18(0.18) NS	0.05(0.09) NS	0.00(0.04) NS	0.12(0.09) NS
Waist	Selected	**GKC**	0.16(0.04)	0.36(0.06)			0.20(0.03)
	Full	**GKFSC**	0.15(0.04)	0.44(0.17) NS	0.00(0.09) NS	0.00(0.04) NS	0.24(0.09)
WHR	Selected	**GKC**	0.15(0.04)	0.19(0.06)			0.09(0.03)
	Full	**GKFSC**	0.13(0.04)	0.29(0.17) NS	0.00(0.09) NS	0.00(0.04) NS	0.13(0.09) NS
ABSI	Selected	**GKC**	0.10(0.04)	0.19(0.06)			0.05(0.03)
	Full	**GKFSC**	0.08(0.04)	0.27(0.17) NS	0.00(0.08) NS	0.00(0.04) NS	0.08(0.09) NS

Continued on next page

Cardiometabolic Traits							
Urea	Selected	**GF**	0.13(0.03)		0.10(0.02)		
	Full	**GKFSC**	0.15(0.04)	0.00(0.17) NS	0.08(0.09) NS	0.00(0.05) NS	0.04(0.09)
Creatinine	Selected	**GKSC**	0.24(0.04)	0.45(0.05)		0.07(0.03)	0.16(0.03)
	Full	**GKFSC**	0.24(0.04)	0.39(0.18)	0.03(0.09) NS	0.07(0.04)	0.13(0.09) NS
Glucose	Selected	**GC**	0.19(0.03)				0.05(0.03)
	Full	**GKFSC**	0.19(0.04)	0.00(0.17) NS	0.00(0.09) NS	0.09(0.05) NS	0.05(0.09)
TC	Selected	**GFS**	0.17(0.03)		0.09(0.02)	0.12(0.04)	
	Full	**GKFSC**	0.15(0.04)	0.12(0.18) NS	0.05(0.09) NS	0.12(0.04)	0.02(0.09) NS
HDL	Selected	**GKC**	0.30(0.04)	0.26(0.05)			0.15(0.03)
	Full	**GKFSC**	0.29(0.04)	0.35(0.16) NS	0.00(0.08) NS	0.01(0.04) NS	0.19(0.08) NS
SBP	Selected	**GKC**	0.15(0.04)	0.13(0.06)			0.10(0.03)
	Full	**GKFSC**	0.14(0.04)	0.18(0.18) NS	0.00(0.09) NS	0.08(0.05) NS	0.13(0.09) NS
DBP	Selected	**GC**	0.17(0.03)				0.09(0.03)
	Full	**GKFSC**	0.13(0.04)	0.00(0.18) NS	0.04(0.09) NS	0.03(0.05) NS	0.03(0.09) NS
HR	Selected	**GF**	0.14(0.03)		0.10(0.02)		
	Full	**GKFSC**	0.03(0.04) NA	0.00(0.18) NS	0.10(0.09) NS	0.00(0.05) NS	0.00(0.10) NS

NS Not significant. That variance component is non-significant according to LRT with p-value > 0.05.

NA Not available. Cannot test the significance of that variance component because the failure in the reduced model.

The mean estimates for h_g^2, h_{kin}^2, e_f^2, e_s^2 and e_c^2 across all traits in the full model were 0.18, 0.22, 0.03, 0.03 and 0.11, respectively. However, the majority of estimates for parameters other than h_g^2 obtained using the full model were not significantly different from zero according to either the Wald test or LRT performed and had large standard errors in general. These results suggest that the full model 'GKFSC' may suffer from the inclusion of correlated factors, as foreseen in the simulation study, probably due to a low number of different types of pairwise relationship in GS10K.

Therefore, we utilised a model selection procedure designed to provide more precise estimates of the parameters retained in a more robust and parsimonious final model, where the least significant effects are removed from the model. More details about the selection procedure are given in Material and Methods. We have demonstrated the effectiveness of our model selection procedure by simulation in the previous section and Table S2.6.

As shown in Table 2.2, SNP-associated genetic effects (represented by $\mathbf{GRM_g}$) were retained in the final models for all 16 traits, indicating that all traits examined here are heritable. Regarding variation associated with families, pedigree-associated genetic effects (represented by $\mathbf{GRM_{kin}}$) and nuclear family environmental effects (represented by $\mathbf{ERM_{Family}}$) were retained in the final models for 10 and 4 out of 16 traits respectively. However, in GS10K, the data structure did not allow for both familial effects to be retained together in the final models for any trait. Additionally, the final models for glucose and DBP included neither $\mathbf{GRM_{kin}}$ nor $\mathbf{ERM_{Family}}$, which is consistent with the previous conclusion derived from Table S2.3, suggesting that familial effects may be limited for these traits.

The additional environmental influences of couple environmental effects (represented by $\mathbf{ERM_{Couple}}$) were retained in the final models for 12 out of 16 traits and sibling environmental effects (represented by $\mathbf{ERM_{Sib}}$) only remained for creatinine and TC.

Although the final model varied between traits, the model 'GKC' was most often selected (9 out of 16 traits) in the model selection procedure in GS10K. Therefore, this suggests that the common environment shared by couples, SNP-associated and pedigree-associated genetic effects are important for the control of a large proportion

of the human complex traits we examined, while the shared family and full-sibling environment have a more limited impact

SNP-associated genetic effects ($\mathbf{GRM_g}$) in the final models provided estimates of h_g^2 ranging between 0.10 and 0.30 with a mean of 0.19 for the 15 traits, excepting height for which nearly half of its phenotypic variation (0.47) was SNP-associated.

For the 10 traits that retained pedigree-associated genetic effects ($\mathbf{GRM_{kin}}$) in the final models, the estimates of h_{kin}^2 ranged from 0.13 to 0.36 with a mean of 0.26, except for creatinine for which nearly half of its phenotypic variation (0.45) was pedigree-associated. For the 10 traits that retained both $\mathbf{GRM_g}$ and $\mathbf{GRM_{kin}}$ in the final models, the estimates of h_{kin}^2 accounted for 56% of the total heritability ($h_{gkin}^2 = h_g^2 + h_{kin}^2$).

Regarding nuclear family environmental effects, the estimates of e_f^2 for 4 traits that retained $\mathbf{ERM_{Family}}$ in the final models were of 18% for anthropometric and of 10% for cardiometabolic traits.

Creatinine and TC were the only two traits for which the common sibling environment ($\mathbf{ERM_{Sib}}$) was kept in the final models, and e_s^2 contributed 7% and 12% of their phenotypic variance respectively.

For those 12 traits that demonstrated evidence of couple effects (i.e. retained $\mathbf{ERM_{Couple}}$ in the final models), e_c^2 accounted for 13.5% of the phenotypic variance on average (of 15% for anthropometric traits and of 11% for cardiometabolic traits).

Compared to the results from the full model in Table 2.2, using the selected final models provided similar but more precise (i.e. with smaller standard errors) parameter estimates. Therefore, whereas the full models gave a general picture of the important components in the architecture of the traits, the final selected models provided a parsimonious model with more precise estimates of the most important effects.

2.2.5.6 Results for Model Selection and Corresponding Variance Component Estimates in GS20K Analyses

We added an extra 10,000 genotyped and phenotyped individuals from the same population, providing 20,000 individuals in total, in order to confirm and build upon the results of the model selection in a more complex data set. The difference in sample sizes and numbers of different relationships between GS10K and GS20K is shown in Table 2.1. The extra 10,000 genotyped individuals in GS20K consisted mainly of the relatives of those already genotyped in GS10K, which substantially increased the proportion of 2^{nd} and 3^{rd} degree and sibling relationships in GS20K. We repeated the model selection procedure and corresponding variance component analyses using selected models in GS20K to identify changes resulting from the increased complexity and sample size of the population.

Results for model selection and variance component analyses using the final selected model as well as the full model are shown in Table 2.3. In general, the parameter estimates obtained from the full model in GS20K were similar to those obtained from the full model in GS10K but the number of non-significant estimates were much lower due to smaller standard errors. Note that standard errors of estimates are not only reduced using GS20K, but, unlike results from GS10K in Table 2.2, are also similar between full and reduced models, suggesting the change is due to improved structure of the data to separate effects as well as increased sample size.

The final models selected from model selection in GS20K were generally similar to those in GS10K, but, owing to the presence of more nuclear family members and siblings in GS20K, we now had better power to detect the past environmental effects (either nuclear family environment or sibling environment), although the estimated effects were usually small.

Table 2.3 Results of variance component analyses for anthropometric and cardiometabolic traits using final models selected from the stepwise model selection and the full model in GS20K

Trait	Model		GRM_g h^2_g (s.e.)	GRM_{kin} h^2_{kin} (s.e.)	ERM_{Family} e_f^2 (s.e.)	ERM_{Sib} e_s^2 (s.e.)	ERM_{Couple} e_c^2 (s.e.)
Anthropometric Traits							
Height	Selected	**GKFC**	0.43(0.02)	0.45(0.04)	0.01(0.02)		0.12(0.02)
	Full	**GKFSC**	0.43(0.02)	0.44(0.04)	0.01(0.02)	$0.01(0.01)^{NS}$	$0.11(0.02)^{NS}$
Weight	Selected	**GKFC**	0.27(0.02)	0.27(0.05)	0.05(0.02)		0.13(0.03)
	Full	**GKFSC**	0.27(0.02)	0.27(0.05)	$0.04(0.02)^{NS}$	$0.02(0.01)^{NS}$	0.13(0.03)
Fat	Selected	**GKSC**	0.24(0.02)	0.25(0.05)		0.04(0.01)	0.19(0.02)
	Full	**GKFSC**	0.24(0.02)	0.22(0.05)	$0.02(0.02)^{NS}$	0.03(0.01)	0.17(0.03)
BMI	Selected	**GKFC**	0.25(0.02)	0.23(0.05)	0.05(0.02)		0.15(0.03)
	Full	**GKFSC**	0.25(0.02)	0.23(0.05)	0.04(0.02)	$0.01(0.01)^{NS}$	0.15(0.03)
Hips	Selected	**GKFC**	0.22(0.02)	0.20(0.05)	0.05(0.02)		0.12(0.03)
	Full	**GKFSC**	0.22(0.02)	0.20(0.05)	0.05(0.03)	$0.01(0.01)^{NS}$	0.12(0.03)
Waist	Selected	**GKSC**	0.19(0.02)	0.31(0.03)		0.04(0.01)	0.18(0.02)
	Full	**GKFSC**	0.19(0.02)	0.25(0.05)	$0.03(0.02)^{NS}$	0.03(0.01)	0.15(0.03)
WHR	Selected	**GKSC**	0.11(0.02)	0.19(0.03)		0.04(0.02)	0.08(0.03)
	Full	**GKFSC**	0.11(0.02)	0.22(0.05)	$0.00(0.03)^{NS}$	0.03(0.02)	0.09(0.03)
ABSI	Selected	**GKSC**	0.11(0.02)	0.21(0.03)		0.03(0.02)	0.05(0.03)
	Full	**GKFSC**	0.10(0.02)	0.24(0.05)	$0.00(0.03)^{NS}$	$0.02(0.02)^{NS}$	$0.07(0.03)^{NS}$

Continued on next page

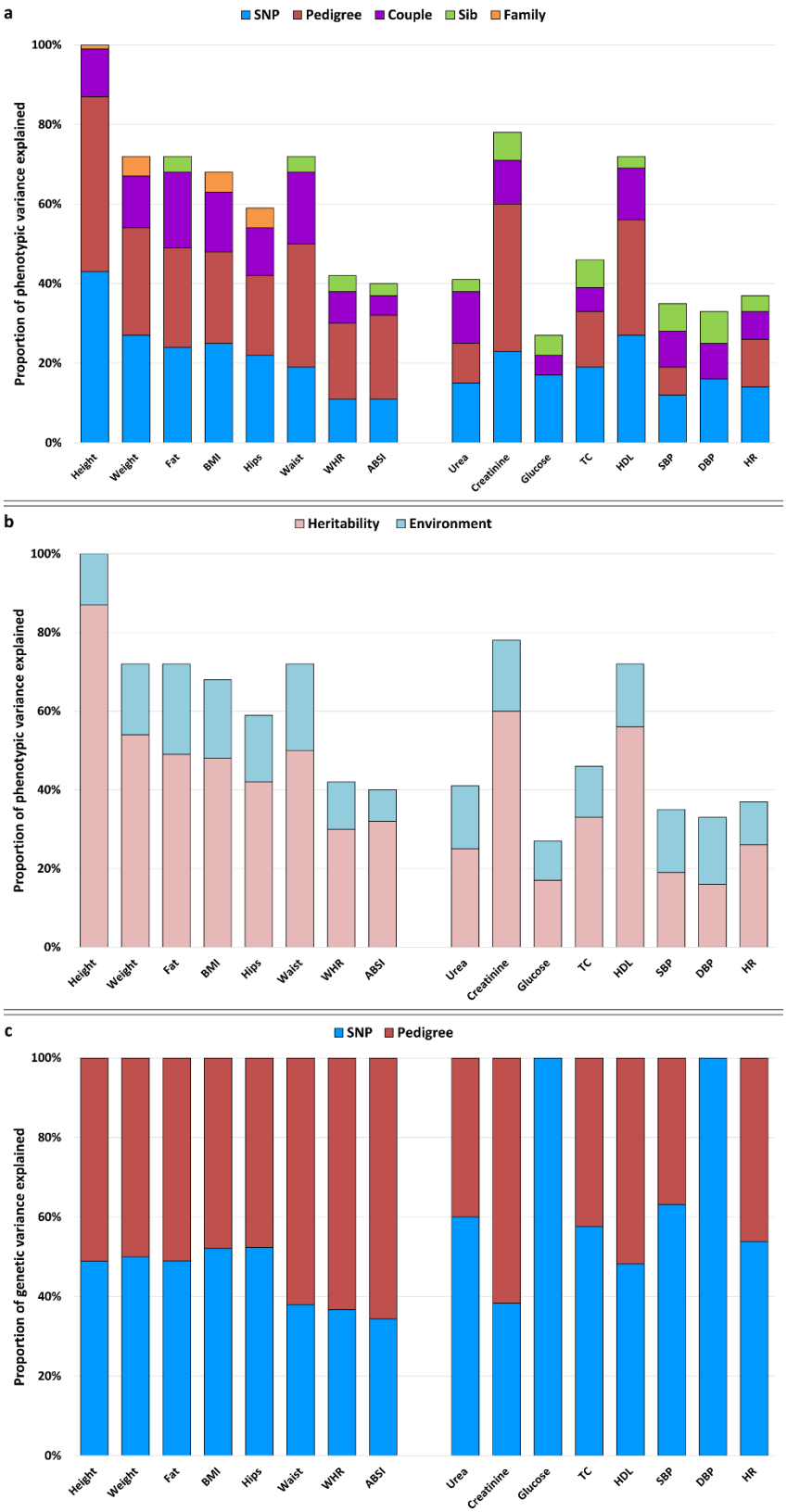
Cardiometabolic Traits							
Urea	Selected	**GKSC**	0.15(0.02)	0.10(0.03)		0.03(0.02)	0.13(0.03)
	Full	**GKFSC**	0.14(0.02)	0.16(0.05)^{NS}	0.00(0.03)^{NS}	0.03(0.02)^{NS}	0.16(0.03)
Creatinine	Selected	**GKSC**	0.23(0.02)	0.37(0.03)		0.07(0.01)	0.11(0.02)
	Full	**GKFSC**	0.21(0.02)	0.46(0.05)	0.00(0.02)^{NS}	0.07(0.01)	0.14(0.03)^{NS}
Glucose	Selected	**GSC**	0.17(0.02)			0.05(0.01)	0.05(0.02)
	Full	**GKFSC**	0.15(0.02)	0.00(0.05)^{NS}	0.03(0.03)^{NS}	0.04(0.02)	0.03(0.03)^{NS}
TC	Selected	**GKSC**	0.19(0.02)	0.14(0.03)		0.07(0.02)	0.06(0.02)
	Full	**GKFSC**	0.19(0.02)	0.12(0.06)	0.01(0.03)^{NS}	0.06(0.02)	0.05(0.03)^{NS}
HDL	Selected	**GKSC**	0.27(0.02)	0.29(0.03)		0.03(0.01)	0.13(0.02)
	Full	**GKFSC**	0.27(0.02)	0.27(0.05)	0.01(0.02)^{NS}	0.02(0.01)	0.11(0.03)
SBP	Selected	**GKSC**	0.12(0.02)	0.07(0.03)		0.07(0.02)	0.09(0.02)
	Full	**GKFSC**	0.12(0.02)	0.08(0.05)^{NS}	0.00(0.03)^{NS}	0.07(0.02)	0.09(0.03)
DBP	Selected	**GSC**	0.16(0.02)			0.08(0.01)	0.09(0.02)
	Full	**GKFSC**	0.14(0.02)	0.00(0.05)^{NS}	0.02(0.03)^{NS}	0.06(0.02)	0.07(0.03)
HR	Selected	**GKSC**	0.14(0.02)	0.12(0.03)		0.04(0.02)	0.07(0.02)
	Full	**GKFSC**	0.14(0.02)	0.10(0.05)	0.01(0.03)^{NS}	0.04(0.02)	0.06(0.03)

^{NS} Not significant. That variance component is non-significant according to LRT with p-value > 0.05.

Moreover, due to an increased number and higher proportion of 2nd and 3rd degree relatives, we had better resolution for familial effects in GS20K. Pedigree-associated genetics and nuclear family environment were now separable and the data structure in GS20K can provide sufficient evidence for both types of familial effects. For weight, urea, TC and HR, familial effects switched from nuclear family environment in GS10K to pedigree-associated genetics or pedigree-associated genetics plus nuclear family environment in GS20K. However, as in GS10K (Table 2.2 and Table S2.3), there was still no evidence of either genetic or environmental familial effects for glucose and DBP in GS20K. The results from final selected models in GS20K are summarized in Figure 2.4.

The heritability estimate is nearly 90%, 60% and 60% for height, creatinine and HDL respectively, and for the remaining anthropometric and cardiometabolic traits, it ranges from 30%-50% and 20-30% for the two types of trait, respectively (Figure 2.4b). Although the proportion of genetic variance explained by SNP-associated and pedigree-associated genetic effects varies across traits, each genetic effect explains around 50% of the genetic variance on average (Figure 2.4c). In GS20K, the most commonly selected model was 'GKCS' (10 out of 16 times, Figure 2.4a and Table 2.3). SNP-associated genetic effects, pedigree-associated genetic effects, sibling environment and couple environment appeared in the final models for 16, 14, 12 and 16 out of 16 times respectively and the means of estimates for h_g^2, h_{kin}^2, e_s^2 and e_c^2 for traits which retained corresponding matrices ($\mathbf{GRM_g}$, $\mathbf{GRM_{kin}}$, $\mathbf{ERM_{Sib}}$ and $\mathbf{ERM_{Couple}}$ respectively) in the final models were of 0.20, 0.23, 0.05 and 0.11 respectively (Figure 2.4a and Table 2.3). For the nuclear family environment, the mean of estimates for e_f^2 for 4 traits which retained $\mathbf{ERM_{Family}}$ in final models was of 0.04 (Figure 2.4a and Table 2.3). On average across traits, our environmental matrices and the final selected models retained through our model selection procedure could explain ~16% and ~56% of the total phenotypic variance respectively (Figure 2.4b).

Figure 2.4 Results of variance component analysis using final selected models for anthropometric and cardiometabolic traits in GS20K.



X-axis:

names of phenotype;

Y-axis:

proportion of phenotypic/genetic variance explained by the different components.

a) Proportion of phenotypic variance explained by genetics and environment for each trait.

b) Proportion of phenotypic variance explained by different components kept in the selected model for each trait.

The major change in GS20K compared to GS10K is the significant evidence of effects of the sibling environment, particularly for cardiometabolic traits, resulting from the higher proportion of sibling relationships in GS20K (more than 12 times compared to GS10K, Table 2.1). However, the sibling effects were only 5% on average and were still relatively low compared to genetic effects and couple environment. Therefore, despite the change in population structure in GS20K, the major components for anthropometric and cardiometabolic traits were SNP-associated and pedigree-associated genetic effects and couple environment as they were in GS10K (Table 2.2).

2.2.6 Discussion

The aim of this study was to better understand the architecture of human complex traits by dissecting phenotypic variation into SNP-associated additive genetic variation (h_g^2), pedigree-associated genetic variation (h_{kin}^2) and environmental influences of common environment shared by nuclear family members (e_f^2), full-siblings (e_s^2) and couples (e_c^2). We generated five design matrices $\mathbf{GRM_g}$, $\mathbf{GRM_{kin}}$, $\mathbf{ERM_{Family}}$, $\mathbf{ERM_{Sib}}$ and $\mathbf{ERM_{Couple}}$ to describe the five effects and we examined 16 human complex traits using genome-wide genotype data and genealogical information in the Generation Scotland: Scottish Family Health study (GS:SFHS) comprising samples from up to 20,000 individuals.

The results of these analyses suggest that SNP-associated genetic effects, pedigree-associated genetic effects and current environment shared by couples were the major contributors to phenotypic variation for anthropometric and cardiometabolic traits. Past environmental influences, such as shared sibling environment or nuclear family environment, made relatively small or undetectable contributions to trait variation (Table 2.2 and Table 2.3). The relative importance of a couple or spousal effect for most traits was also noted by Liu et al. [15], in analyses based only on pedigree relationships, although they did not find a significant spousal effect for cholesterol, HDL or glucose for which a significant couple effect was detected in this study.

Considering the low number of non-zero off-diagonal entries in $\mathbf{ERM_{Couple}}$ (1,283 or 1,767 pairs in GS10K or GS20K), the signal of couple effects was quite strong. We

did observe significant phenotypic correlation between couple pairs for almost all traits in our data (Table S2.7). For some traits this presumably represents current shared environment due to cohabitation, such as living habits and diet. For traits related to obesity, it is reasonable that current environmental effects are more important than past environmental effects since traits like BMI, fat, HDL and blood pressure are potentially influenced by recent food intake, exercise and medical treatment.

It should be noted that in our sample participants have an average age of ~50 years and individuals currently sharing a common household environment will largely be couples, whereas most individuals involved in sibling and parent-offspring relationships will no longer be cohabiting at the point when the data were recorded. It has been previously reported in obesity studies that common childhood environment only affects individuals in their mid-childhood but the influence does not last past adolescence [23,24]. Therefore, although the impacts of nuclear family or sibling environmental effects on the 16 traits we examined were relatively small, family and sibling environmental effects could be more important in younger cohorts and might be of greater importance for other complex traits and diseases where long-term environment may have an influence on a phenotype that is relatively stable over time.

For some traits, the most obvious example being height, couple effects may also, in part or completely, reflect assortative mating. A study by Keller et al. has shown that h^2 estimate for height would be 13% higher with assortative mating than it would have been under random mating [109]. If there was assortative mating for any of the traits which retained $\mathbf{ERM_{Couple}}$ in final models but we modelled the couple correlation as an environmental effect, we would expect to obtain biased e_c^2 estimates. Moreover, modelling assortative mating as an environmental effect removes variance from the residual ("error") variance. We therefore might obtain an inflated h_g^2 estimate if we have not taken assortative matting into account and reduce the residual variance as a consequence of modelling assortative matting as an environmental effect. In addition, assortative mating will have consequences for our interpretation of GWAS results as the combined effect of detected loci on the trait variance will be greater than the sum of the effects of the individual loci due to the positive correlations between loci. However, except for height, where the phenotype will be largely fixed by the time of

marriage, for most traits it is difficult to determine whether assortative mating and/or shared environment are responsible for observed phenotypic correlations between couples.

Shared sibling environment was undetected for most of the traits in GS10K (Table 2.2), whereas there was significant evidence of it for many traits in GS20K (Table 2.3), indicating that the detection power of sibling environment benefits from the increase in number and proportion of sibling relationships (Table 2.1). Sibling effects, where detected, explained 5%, on average, of the trait variation. Estimated sibling effects may be inflated by non-additive genetics, (i.e. dominance and epistasis). As sibling effects only capture a fraction of the non-additive variation, the actual variation contributed by non-additive genetics might potentially be large and would merit further study.

Our analyses split the genetic variation approximately equally on average across traits between that which was associated with SNPs (h_g^2) and that which was associated with pedigree (h_{kin}^2). A plausible interpretation for the division of genetic effects into h_g^2 and h_{kin}^2 is that h_g^2 is able to explain the genetic variation attributed by common variants inherited from distant ancestors that are in LD at the population level and are well captured due to association with genotyped SNPs [76]. On the other hand, h_{kin}^2 accounts for the genetic variation due to rare variants, CNVs and other structural variation, etc. that cluster in specific families and are captured due to strong linkage in high-order pedigrees but are not in population-wide LD with common SNPs.

We compared h_g^2 and h_{gkin}^2 (calculated as $h_g^2 + h_{kin}^2$) estimates obtained in final models from model selection in GS20K to two relevant publications from Zaitlen et al. [76] and Vattikuti et al. [107] that also explored the influence of including relatives on h^2 estimation in family-based studies and compared h_{gkin}^2 estimates obtained in final models in GS20K to published twin studies [73,110-117]. Comparisons are shown in Table 2.4.

Table 2.4 Comparisons of the results from final models in GS20K to previous published results

Family-based GREML Studies				
Trait	**Final models**		**Publications**	
	$h_g^2(s.e.)$	$h_{gkin}^2(s.e.)$	$h_g^2(s.e.)$	h_{gkin}^2/h_{ped}^2[a] $(s.e.)$
Height	0.43(0.02)	0.88(0.03)	0.40 [76]	0.69 [76]
BMI	0.25(0.02)	0.48(0.04)	0.14-0.23 [76,107]	0.34-0.42 [76,107]
WHR	0.11(0.02)	0.30(0.03)	0.06-0.13 [76,107]	0.19-0.28 [76,107]
Glucose	0.17(0.02)	0.17(0.02)	0.10 [107]	0.33 [107]
HDL	0.27(0.02)	0.56(0.03)	0.12-0.24 [76,107]	0.45-0.48 [76,107]
SBP	0.12(0.02)	0.19(0.03)	0.24 [107]	0.30 [107]

Twin Studies		
Trait	**Final models**	**Publications**
	h_{gkin}^2 $(s.e.)$	h_{ped}^2 $(s.e.)$
Height	0.88(0.03)	0.89 - 0.93 [73]
Weight	0.54(0.04)	0.64-0.84 [110]
Fat	0.49(0.04)	0.59-0.63 [111]
BMI	0.48(0.04)	0.48-0.61 [112]
Hips	0.42(0.04)	0.52-0.58 [111]
Waist	0.50(0.03)	0.46 [113]
WHR	0.30(0.03)	0.31 [113]
Urea	0.25(0.03)	0.36-0.54 [115]
Creatinine	0.60(0.03)	0.37 [114]
Glucose	0.17(0.02)	0.45 [116]
TC	0.33(0.03)	0.46-0.57 [112]
HDL	0.56(0.03)	0.50-0.62 [112]
SBP	0.19(0.03)	0.57 [117]
DBP	0.16(0.02)	0.45 [117]
HR	0.26(0.03)	0.64 [117]

[a] h_{gkin}^2 is an equivalent estimate to h_{ped}^2 but is calculated using genomic information

When comparing with two family-based GREML studies (Table 2.4), our h_g^2 and h_{gkin}^2 estimates from final models are generally higher than published relevant results, except for the h_g^2 estimate for SBP and the h_{gkin}^2 estimates for glucose and SBP. When comparing with twin studies (Table 2.4), our h_{gkin}^2 estimates for all anthropometric traits, urea, TC and HDL given by final selected models in GS20K are reasonably close to reported h_{ped}^2 estimates, which suggests little missing heritability. Hence, our results provide no evidence that heritabilities given by previous twin studies were inflated for these traits. For glucose, SBP, DBP and HR, however, our h_{gkin}^2 estimates are significantly lower than previously published estimates of h_{ped}^2, whereas for creatinine, h_{gkin}^2 is significantly larger.

To validate the analytical approach used in this study and to evaluate model robustness, we conducted a detailed simulation study using real genotype and pedigree information obtained from GS10K. The simulation results confirmed that our models were generally robust (Table S2.5). However, the inevitable correlations between our design matrices can, under some circumstances, make it challenging to partition variance for correlated factors in variance component analyses and accurately discriminate between competing models in model selection. Nonetheless, any influence of inaccurately partitioning variance among correlated matrices was relatively limited and our models were always able to provide us with a good idea of the magnitude of corresponding effects as the mean estimate for each parameter was always very close the simulated settings when the model used for analysis matched the simulated sources of trait variation.

The effectiveness of the model selection procedure was also validated using the simulated data with the model selection procedure often ($\geqslant$80%) resulting in models containing all major phenotype components (Table S2.6). However, due to the limited number of appropriate relationships in GS10K to resolve correlations between matrices and to detect factors with small effects, our model selection procedure may omit minor effects (contributing 5% or less of the trait variance, for example). In addition, the procedure may sometimes identify incorrect models (not being able to distinguish familial effects as mentioned in the simulation study and Table S2.6) and

this might be the case for weight, urea, TC and HR in Table 2.2. However, with sufficient data from higher order pedigree relationships, as was the case in GS20K, the impact of covariances between design matrices in first order relatives (parent-offspring, siblings and couples) are mitigated and further components of variance became separable (Table 2.3).

To sum up, we provide evidence that for the traits we have analysed, heritabilities are divided approximately evenly between pedigree-associated and SNP-associated genetic effects. This is the case even when, as here, we have taken care to consider various models of environmental covariation of first-degree relatives (including couples). It appears that confounding factors like dominance, shared full-sibling environment and the past rearing environment seem to have relatively small contribution to phenotypic variation for these traits in our population. We find that current shared environment of couples is able to account for another ~11% on average of the phenotypic variation of human complex traits. This has been seldom mentioned in previous heritability studies but we note that as an effect that inflates the covariance between nominally unrelated individuals, it should not substantially bias or inflate h^2_{ped} and h^2_{gkin}. It should be taken into account that couple effects may also be present in cohorts of unrelated individuals which may often include couples but ignore any correlation between them. Therefore, it might bias h^2_g from genotype-based studies which do not account for such couple effects and could have an impact on GWAS studies.

Overall, our work shows that SNP-associated genetic effects, pedigree-associated genetic effects and current shared couple environmental effects are three fundamental components of phenotypic variation for traits related to anthropometrics and cardiometabolism and current shared environmental effects have more impact than past shared environmental effects. This also has implications for models to be used in further studies of the architecture of complex traits including utilising the appropriate models for GWAS and related analyses and for personalised disease risk prediction.

2.2.7 Material and Methods

2.2.7.1 Ethics Statement

Ethical approval for the study was given by the NHS Tayside committee on research ethics (reference 05/s1401/89). Governance of the study, including public engagement, protocol development and access arrangements, was overseen by an independent advisory board, established by the Scottish government.

2.2.7.2 Data Description

Our dataset came from the Generation Scotland Scottish Family Health Study (GS:SFHS) project (http://www.generationscotland.org), which was collected by a cross-disciplinary collaboration of Scottish medical schools and the National Health Service (NHS) from Feb 2006 to Mar 2011 [118,119].

Data for 16 complex traits were used. These were 8 anthropometric traits: height, weight, fat, body mass index (BMI $= \frac{\text{Weight}}{\text{Height}^2}$), hips, waist, waist-to-hips ratio (WHR) and a body shape index (ABSI $= \frac{\text{Waist Circumference} \times \text{Height}^{5/6}}{\text{Weight}^{2/3}}$) [108] and 8 cardiometabolic traits: levels of creatinine, urea, total cholesterol (TC) and high density lipoprotein (HDL) in serum and glucose in blood after a four hour fast period, systolic blood pressure (SBP), diastolic blood pressure (DBP) and heart rate (HR). None of the traits was adjusted for medication or fasting status. We explored the phenotypic distributions of these traits and conducted natural logarithm transformations for them, except for height, sodium and fat, to obtain approximate normal distributions. We set phenotypes with values greater or smaller than the mean ± 4 standard deviations (after adjusting for sex, age and age^2) to missing.

Data also contained the information of sex, age, clinics where the phenotypes were measured and Scottish Index of Multiple Deprivation (SIMD, an environmental ranking based on living areas, [120]). A descriptive analysis can be seen in Table S2.1.

The first set of analyses presented in the manuscript are based on a data set of nearly 10,000 individuals from GS:SFHS (GS10K). These have multiple degrees of kinships,

including 5,061 family members from 1,612 nuclear or extended families, and were genotyped with the Illumina OMNiExpress chip (707,686 SNPs). We conducted data quality control in Plink v1.07 [121] and GenABEL v1.7-6 [122]. SNPs with a minor allele frequency (MAF) < 0.05, a Hardy-Weinberg Equilibrium's (HWE) p-value $<10^{-6}$ and a missingness > 2% were excluded. Duplicate samples, gender discrepancies and individuals with more than 5% missingness were also removed. After the quality control we kept 9,863 individuals genotyped for 550,796 common SNPs over the 22 autosomes.

An extended dataset (GS20K) was used to validate the results obtained with GS10K and evaluate the effect of including further close relationships in our data. The extra 10,000 individuals were genotyped with the same chip and quality control was performed using the same criteria as in the GS10K. After quality control, GS20K consisted of 20,032 individuals, 18,293 of whom came from 6,578 nuclear or extended families, and 519,729 common SNPs across the 22 autosomes.

A comparison of the difference in relationships between GS10K and GS20K can be seen in Table 2.1.

2.2.7.3 Statistical Methods

Our model allows trait variation to be influenced by the genetic effects associated with SNPs (h_g^2) and pedigree (h_{kin}^2) and the environmental effects shared by families (e_f^2), couples (e_c^2) and full-siblings (e_s^2), (Figure 2.1). To estimate the influence of each effect, we generated five design matrices: $\mathbf{GRM_g}$, $\mathbf{GRM_{kin}}$, $\mathbf{ERM_{Family}}$, $\mathbf{ERM_{Sib}}$ and $\mathbf{ERM_{Couple}}$.

2.2.7.3.1 Genomic Relationship Matrices

A genomic relationship matrix (GRM) contains estimated genomic relatedness between pairs of individuals calculated from identity-by-state marker relationships as in Yang et al. [51,52].

Each off-diagonal entry in the GRM represents the realised genomic relationship between a pair of individuals:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{(x_{ji}-2p_i)(x_{ki}-2p_i)}{2p_i(1-p_i)}$$

where, p_i is the minor allele frequency (MAF) for SNP i, x_{ji} or x_{ki} is the allelic dose for individual j or k at locus i (x = 2 if the individual carries two rare alleles, x = 1 if the individual is heterozygous, x = 0 if the individual carries two common alleles) and N is the total number of SNPs.

Each entry on the diagonal represents the inbreeding coefficient calculated as:

$$1+\frac{1}{N}\sum_{i=1}^{N}\frac{x_{ji}^2-(1+2p_i)x_{ji}+2p_i^2}{2p_i(1-p_i)}$$

We used GCTA [52] to generate $\mathbf{GRM_g}$ and obtained $\mathbf{GRM_{kin}}$ by modification of $\mathbf{GRM_g}$ in R [123]. Their definitions are identical to matrices $\mathbf{K_{IBS}}$ and $\mathbf{K_{IBS>}}_t$ in Zaitlen et al. [76] respectively.

$\mathbf{GRM_g}$: a GRM estimated using all common SNPs, and designed to capture the additive genetic variance explained by common SNPs in the population sample.

$\mathbf{GRM_{kin}}$: a modified GRM calculated as in Zaitlen et al. [76] designed to estimate the extra genetic effects associated with pedigree, the variance explained by shared genetic factors in close relatives. $\mathbf{GRM_{kin}}$ was created by setting to 0 all entries in $\mathbf{GRM_g}$ smaller than 0.025.

The number of entries different from 0 in each of the matrices is shown in Table 2.1.

2.2.7.3.2 Environmental Relationship Matrices

An environmental relationship matrix (ERM) is a covariance matrix designed to capture the variance due to common environmental effects shared among a specified group of individuals.

The ERM coefficient for each pair of individuals is 1 in if they share a particular environment, e.g., living in the same area or coming from the same family; otherwise, it is 0. Each entry on the diagonal is 1.

We generated 3 different ERMs in R [123]: $\mathbf{ERM_{Couple}}$, $\mathbf{ERM_{Sib}}$ and $\mathbf{ERM_{Family}}$.

$\mathbf{ERM_{Couple}}$: $\mathbf{ERM_{Couple}}$ was designed to capture the common environmental effects shared between a couple. The ERM coefficient of two individuals was 1 if they were identified as a couple, defined as a pair of individuals with at least one offspring within GS:SFHS. Each entry on the diagonal was 1.

$\mathbf{ERM_{Sib}}$: $\mathbf{ERM_{Sib}}$ was designed to capture the common environmental effects shared between full-siblings. The ERM coefficient of two individuals was 1 if they were identified as full-siblings. Each diagonal entry was 1.

$\mathbf{ERM_{Family}}$: $\mathbf{ERM_{Family}}$ was designed to capture the common environmental effects shared within each nuclear family comprising parents and offspring. The ERM coefficient of two individuals was 1 if they were identified as a parent-offspring pair, full-siblings or a couple. The ERM coefficient of two individuals was 1 if they were identified as nuclear family members, including parent-offspring, couple and full-sibling relationships. Each diagonal entry was 1.

The number of entries different from 0 in each of the environmental matrices is shown in Table 2.1. Details about model and matrices we defined can be seen in Figure 2.1.

2.2.7.3.3 Variance component analysis

We used the genomic and environmental matrices described above to partition the phenotypic variance observed for the traits using a mixed model in a restricted maximum likelihood (REML) framework. The analyses were implemented in GCTA [52]. The equations used to evaluate each model were the subsets of the full model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{g_g} + \mathbf{g_{kin}} + \mathbf{e_f} + \mathbf{e_s} + \mathbf{e_c} + \boldsymbol{\varepsilon}\text{, with}$$

$$\mathbf{V} = \mathbf{GRM_g}\sigma_g^2 + \mathbf{GRM_{kin}}\sigma_{kin}^2 + \mathbf{ERM_{Family}}\sigma_{ef}^2 + \mathbf{ERM_{Sib}}\sigma_{es}^2 + \mathbf{ERM_{Couple}}\sigma_{ec}^2 + \mathbf{I}\sigma_\varepsilon^2$$

where $\mathbf{y}$ is an $n \times 1$ vector of observed phenotypes with n being the sample size (number of individuals), and $\mathbf{V}$ the total phenotypic variance matrix, $\boldsymbol{\beta}$ is an $m \times 1$ vector of fixed effects with m being the total level of covariates and $\mathbf{X}$ its design matrix with dimension $n \times m$, $\mathbf{g_g}$ is an $n \times 1$ vector of the total additive genetic effects of the individuals captured by genotyped SNPs with $\mathbf{g_g} \sim N(0, \mathbf{GRM_g}\sigma_g^2)$, $\mathbf{g_{kin}}$ is an $n \times 1$ vector of the extra genetic effects associated with the pedigree for relatives with $\mathbf{g_{kin}} \sim N(0, \mathbf{GRM_{kin}}\ \sigma_{kin}^2)$, $\mathbf{e_f}$, $\mathbf{e_s}$ and $\mathbf{e_c}$ are $n \times 1$ vectors representing the common environmental effects shared by nuclear family members, full-siblings and couples with $\mathbf{e_f} \sim N(0, \mathbf{ERM_{Family}}\ \sigma_{ef}^2)$, $\mathbf{e_s} \sim N(0, \mathbf{ERM_{Sib}}\ \sigma_{es}^2)$ and $\mathbf{e_c} \sim N(0, \mathbf{ERM_{Couple}}\ \sigma_{ec}^2)$ and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of residuals. We fitted a range of models including different combinations of effects, and named them using abbreviations according to the effects used. We used the codes 'G' for $\mathbf{GRM_g}$, 'K' for $\mathbf{GRM_{kin}}$, 'F' for $\mathbf{ERM_{Family}}$, 'S' for $\mathbf{ERM_{Sib}}$ and 'C' for $\mathbf{ERM_{Couple}}$ –e.g. model 'GKC'$= \mathbf{GRM_g} + \mathbf{GRM_{kin}} + \mathbf{ERM_{Couple}}$, and the proportion of total phenotypic variance captured by each matrix was termed h_g^2, h_{kin}^2, e_f^2, e_s^2 and e_c^2 accordingly. All models include a residual matrix and the total heritability h_{gkin}^2 is always the sum of $h_g^2 + h_{kin}^2$ for any model.

There were 31 different models from all the possible combinations of the five matrices. The abbreviations for each model and the formulae to estimate each term in each model are listed in Table S2.2. The results for each model are listed in Table S2.4.

In addition to the matrices described (including the residual matrix), we always included the fixed effects of sex, age, age^2, sex-by-age interaction, clinic, standardised SIMD and SIMD^2 and the first 20 eigenvectors of $\mathbf{GRM_g}$ (to ameliorate problems associated with data structure).

2.2.7.3.4 Stepwise model selection

We conducted a stepwise model selection to find the most appropriate genetic and environmental model for each trait and dissect the phenotypic variation into its components (SNP-associated additive genetic variance, pedigree-associated genetic effects shared among relatives and common environmental effects shared among the specified groups including nuclear family members, couples and full-siblings).

The stepwise selection began with the full model 'GKFSC', where all matrices were fitted together. We performed a Wald test and a log-likelihood ratio test (LRT, using a mixture distribution of $\chi^2_{df=0}$ and $\chi^2_{df=1}$ with a probability of 0.5 [52]) for each component and removed the component, if any, that was non-significant for both tests at $\alpha = 5\%$ level and had the highest p-value for the Wald test. We repeated this process until all the remaining components were significant for at least one test. We did not correct for the limited number of traits analysed so error rates in this procedure should be considered to be on a per trait basis.

2.2.7.4 Simulation Study

In order to evaluate the robustness of our models and the performance of our stepwise model selection, we conducted a simulation study. We simulated, based on the real genotypic information and the real pedigree, different sets of phenotypes for each of the 9,863 individuals in GS10K.

For simulating the genetic effects, we used a similar approach to Zaitlen et al. [76] by dividing the genome into two: even and odd chromosomes, and randomly selecting 550 SNPs from even and odd chromosomes (approximately 1 from each 500 SNPs), representing the observed causal loci that were in LD with the SNPs (SNP-associated genetic effects) and the unobserved genetic variants that were not in LD with the SNP array (pedigree-associated genetic effects) separately. In a later step, only even chromosomes were used to generate $\mathbf{GRM_g}$ and $\mathbf{GRM_{kin}}$. Each locus was assigned an effect size driven from exponential distribution as in Fisher [124] and the summed effects for even and odd chromosome SNPs were designed to explain h^2_g and h^2_{kin} of the trait variance respectively.

For environmental factors, we simulated a sibling environmental effect, a couple environmental effect and two nuclear family environmental effects (youth and adulthood environments) for each individual. The corresponding effect sizes for sibling, couple and nuclear family environmental effects were derived from $N(0, e^2_s)$, $N(0, e^2_c)$ and $N(0, e^2_f)$ accordingly and were the same among full-siblings, between couples and among nuclear family members.

In addition, we simulated a random residual effect for each individual, the residuals were derived from $N(0, e_e^2)$ where e_e^2 represents the proportion of variance remaining in each of the scenarios. For each scenario, each component ($h_g^2, h_{kin}^2, e_c^2, e_s^2, e_f^2$) was given a proportion of the variance explained and e_e^2 was $1 - h_g^2 - h_{kin}^2 - e_c^2 - e_s^2 - e_f^2$. The final phenotypes would be the sum of these genetic and environmental effects and residuals, and the expected mean and variance of simulated phenotypes were 0 and 1, respectively. More details about how we simulated phenotypes can be found in Text S2.1.

We evaluated the robustness of our models under situations where phenotypes were contributed by i) one of the five effects, ii) SNP-associated genetic effects and one of the familial effects (either pedigree-associated genetic effects or nuclear family environmental effects) and iii) SNP-associated genetic effects, familial effects and other environmental effects. All scenarios included residuals and 50 to 100 replicates were analysed for each scenario. The results of simulations were evaluated using a Z-test, which tested whether the mean estimate for each parameter deviated significantly from its simulated value. Note, it was too time consuming to explore all the possible combinations of models and simulated phenotypes, therefore, we mainly focused on the models that were selected in model selection procedure for the real phenotypes in GS10K (Table 2.2) as well as the fundamental models of our study. More details about the parameter settings for these scenarios can be found in Table S2.5.

$\mathbf{ERM_{Family}}$ posited a relationship between siblings, parents-offspring and couples is somewhat confounded with the addition of $\mathbf{GRM_{kin}}$ and $\mathbf{ERM_{Couple}}$, making separation and estimation of these effects (e_f^2, h_{kin}^2 and e_c^2) challenging, as confirmed by the results from analysis of real phenotypes in GS10K (Table 2.2). Hence, we evaluated the effectiveness of our model selection procedure under situations where phenotypes were contributed by moderate SNP-associated genetic effects and low sibling environmental effects plus a) moderate nuclear family environmental effects but low pedigree-associated genetic effects and couple environmental effects, b) low nuclear family environmental effects but moderate pedigree-associated genetic effects and couple environmental effects and c) moderate nuclear family environmental effects, pedigree-associated genetic effects and couple environmental effects. All

scenarios included residuals. More details about the parameter settings for these scenarios can be found in Table S2.6. We conducted the model selection procedure for each replicate to see whether the final model selected matched the simulated phenotypic components for these scenarios (Note: we ran 10 replicates for each scenario here). In addition, variance component analyses were performed using final selected models for these replicates to see whether the estimates of parameters were close to their simulated values.

2.2.8 Acknowledgments

The authors are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland. We are also grateful for comments from three anonymous reviewers that resulted in a much improved manuscript.

2.2.9 Members of Generation Scotland

David Porteous, Blair Smith, Sandosh Padmanabhan, Lynne Hocking, Caroline Hayward, Ian Deary and Andrew McIntosh.

2.3 Conclusion and Discussion

In Chapter 2, I presented my novel method of variance component analysis. Compared to frequently used methods such as GREML and twin studies, my method could: 1) estimate SNP (h_g^2) and total heritability (h_{ped}^2) simultaneously; 2) model environmental contribution to the trait variation; and 3) maximise the sample size.

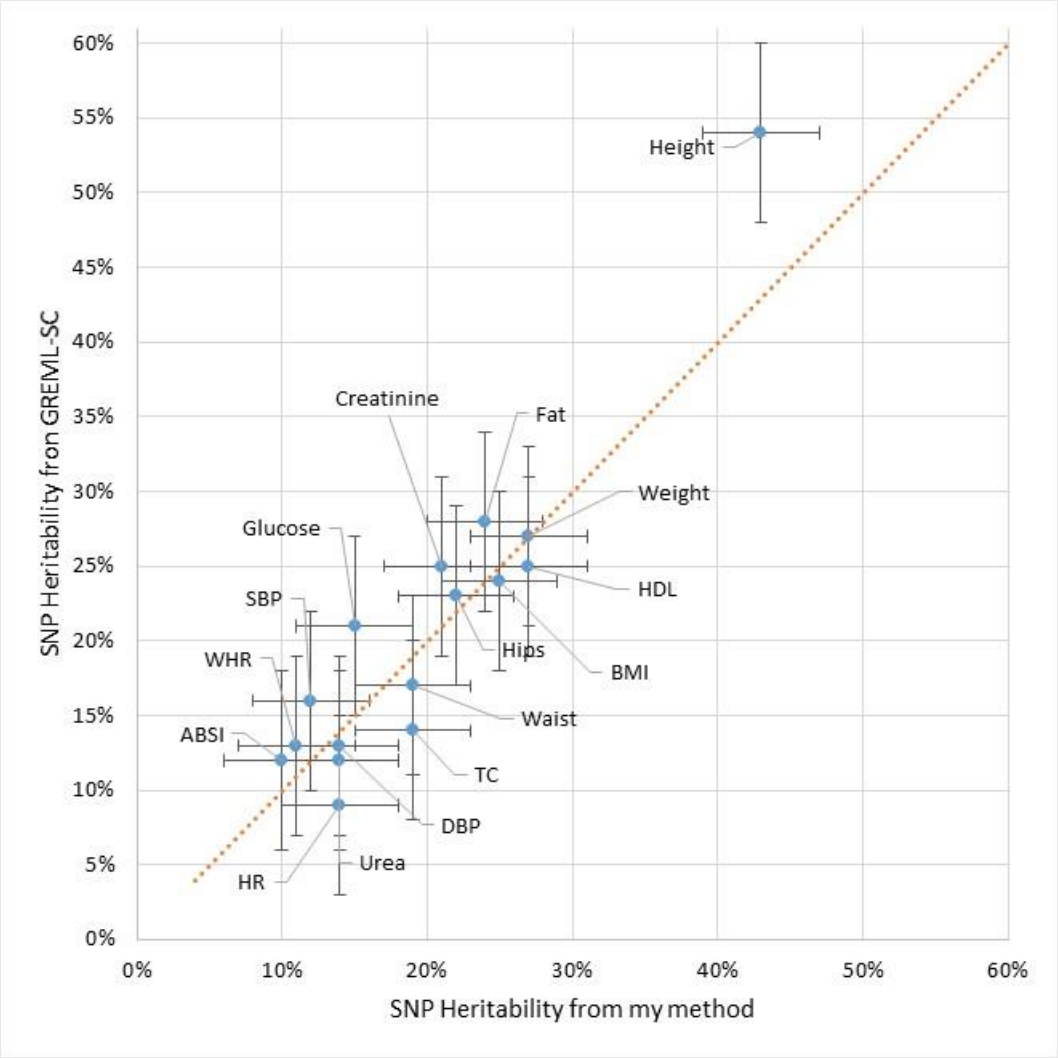
Owing to the sufficiency of multiple different relationships in GS:SFHS (Table 2.1), I was able to disentangle the signal of trait variation from different sources. Simulation study confirms that my method works reasonable well with major components of trait variance usually being detected and the estimation of their contribution being unbiased or only a little biased (1-2% discrepancy between estimates and simulated values). Indeed, completely discriminating familial confounding factors is still challenging and thus systematic bias remains. However, the bias is relatively small and, consequently, the results are reliable.

I investigated 8 anthropometric traits and 8 cardio-metabolic traits; and found that for majority of the traits studied SNP-associated genetic effects, pedigree-associated genetic effects, couple environment and sib environment are the major contributors to trait variation. This reveals new insight into trait components of human complex traits compared to traditional variance component analysis which mainly focuses on the additive genetics.

In my method, the genetic effects are separated into SNP-associated (h_g^2) and pedigree-associated (h_{kin}^2) genetic effects, which I believe represent well-tagged common SNP variants inherited from distant ancestors and untagged variants such as rare SNP variants and non-SNP variants passed from recent ancestors respectively. I compared the estimates of h_g^2 from my method (the proportion of trait variance explained by $\mathbf{GRM_g}$ in the selected model from Table 2.3) and from GREML-SC method [51,52] (Single-component-GREML, which estimates the variance explained by genotyped common SNPs in unrelated individuals, results see Table S2.3 column 1) for the same traits by plotting one against another in Figure 2.5.

Figure 2.5 shows that, for majority of the traits, estimates of h_g^2 from two methods are quite close to one another for the same traits, demonstrating that although relatives were included in the analysis, the confounding factors shared among them did not hinder the method from providing reliable SNP heritability.

Figure 2.5 Comparing the estimates of h_g^2 for the same trait using GREML-SC method and using my method in GS:SFHS.



Y-axis: SNP heritability estimated by GREML-SC; X-axis: SNP heritability estimated selected model using my method; Red dotted line: Y=X; Horizontal and vertical bars: standard errors of the estimates of h_g^2 from my method and GREML-SC method respectively.

I also compared the total heritability estimated from the selected model ($h^2_{gkin} = h^2_g + h^2_{kin}$) with h^2_{ped} from twin studies (Table 2.4) and found that for most of the traits investigated, my h^2_{gkin} estimates are close to published h^2_{ped} estimates which suggest little missing heritability for those traits. This points out that the difference between h^2_{ped} and h^2_g is mainly due to genetic variations contributed by variants that are not tagged by SNP array. But such variants are in strong LD with pedigree and thus could be captured by our method. A similar conclusion was drawn by Yang et al. [125] where they conducted GREML-MS (MAF-stratified-GREML, creating several GRMs using imputed SNPs according to MAF bin and estimating the total variance explained by these GRMs in unrelated individuals) and found that imputed rare SNPs could explain an additional amount of genetic variance, leaving little missing heritability unexplained for height and BMI.

One of the novel findings in this study is I found that current environment shared by partners (members of a couple) are very important to trait variance for almost all the traits I looked at. However, for height, BMI and waist and hip circumference, couple effects may reflect, in part or completely, assortative mating [91,92,109,126]. $\mathbf{ERM_{Couple}}$ is a similarity matrix designed to capture the resemblance between partners. For genetically unrelated couples under random matting pattern, such resemblance should be mainly contributed by shared living environment. However, assortative mating increases both genetic and environmental similarity between partners and thus the presence of assortative mating will inflate the estimates of e^2_c in my method because e^2_c no longer represents the similarity between partners purely due to shared environment but the phenotypic similarity (contributed by both genetics and environment) due to mate choice. I believe this is the reason that there are little or very little residual variance left in our selected model for height. To find out whether the couple effect is due to assortative mating or shared environment or both and how assortative mating influence trait architecture, a follow-up study was conducted in Chapter 5.

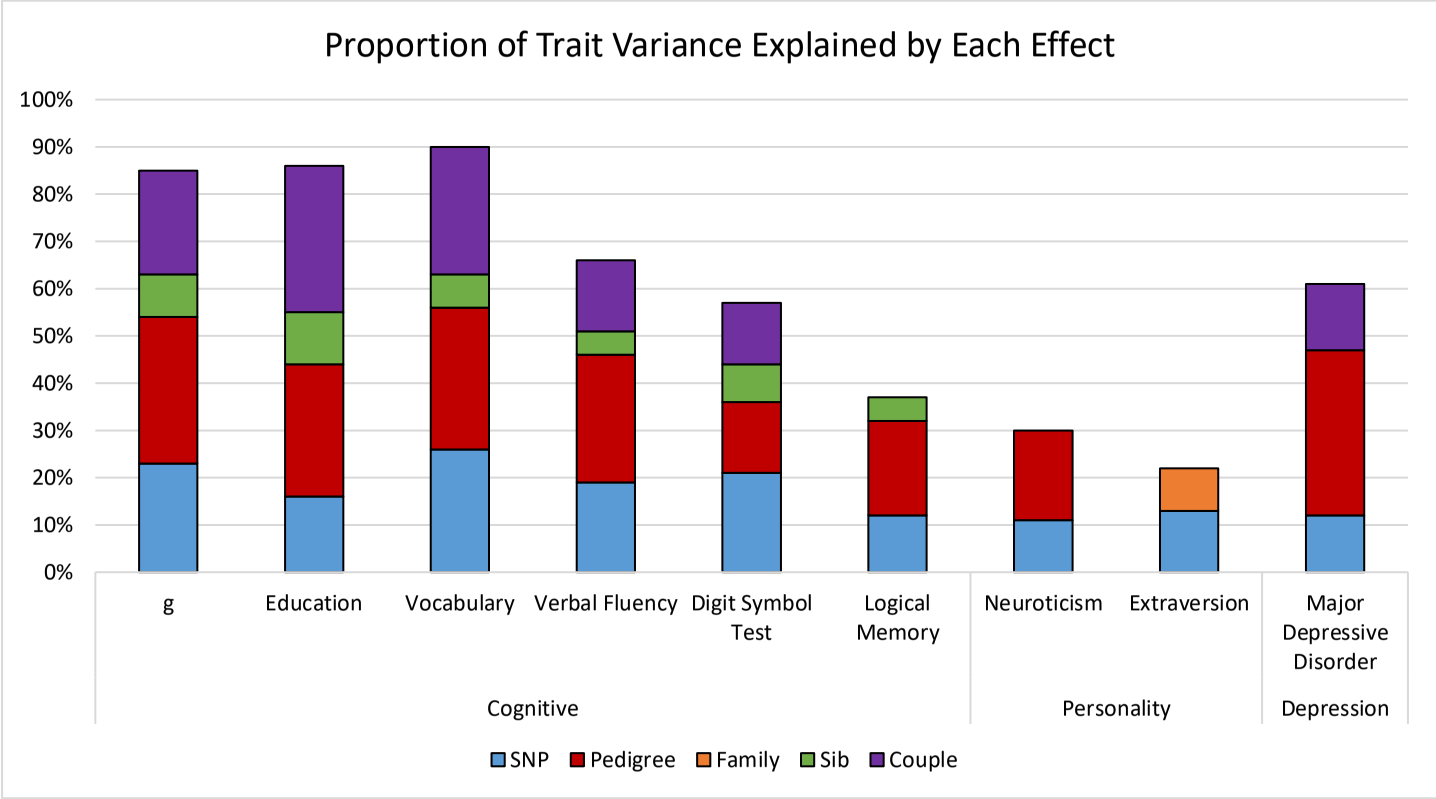
Similarly, $\mathbf{ERM_{Sib}}$ is a similarity matrix designed to capture the resemblance between siblings. Such resemblance could be attributable to additive genetics (which is modelled by two GRMs in the method), non-additive genetics (which is not modelled

in this method) and shared common environment in the past (because most full-siblings no longer lived together when they were recruited). Therefore, the estimates of e_s^2 in this method might be inflated by non-additive genetics such as dominance. However, more recent meta-twin study shows that there is little evidence for dominance for human complex traits [14], i.e. estimates of e_s^2 should be unbiased and reflect the effects of shared rearing environment in the past.

Apart from this research, I have cooperated with colleagues from other departments of my university to study the genetic and environmental contribution to trait variance using this method for major depressive disorder (with Yanni Zeng) [127] and traits related to cognition and personality (with David Hill) [128], including cognitive traits: general intelligence (g), education attainment, vocabulary, verbal fluency, digit symbol test and logical memory; and personality traits: neuroticism and extraversion. The manuscripts of these two papers are attached in the appendix (Publication S2.1 for the former and Publication S2.2 for the later).

The results of the selected models from these two studies are visually plotted in Figure 2.6 and the results of the selected and the full models are recorded in Table S2.8. Regarding cognitive traits, the pattern is quite clear that SNP-associated genetic effects (represented by $\mathbf{GRM_g}$), pedigree-associated genetic effects (represented by $\mathbf{GRM_{kin}}$), couple environmental effects (represented by $\mathbf{ERM_{Couple}}$) and sibling environmental effects (represented by $\mathbf{ERM_{Sib}}$) are four major components of the trait variance, except for logical memory of which the selected model is 'GKS'. The proportion of phenotypic variance contributed by component remaining in the selected model is similar across cognitive traits, which are h_g^2=20%, $h_{kin}^2 = 25\%$, $e_c^2 = 22\%$ and $e_s^2 = 8\%$ on average. For personality traits, the selected models for neuroticism and extraversion are 'GK' (h_g^2=11% and $h_{kin}^2 = 19\%$) and 'GF' (h_g^2=13% and $e_f^2 = 9\%$) respectively, which suggests that their trait architectures are different. Whereas for depression, the selected model for major depressive disorder is 'GKC' with $h_g^2 = 12\%$, $h_{kin}^2 = 35\%$ and $e_c^2 = 14\%$. However, assortative mating has been reported for major depression disorders [129], personality [130] and intelligence [131] and thus the estimates of e_c^2 might represent (or be inflated by) the phenotypic correlation generated by mate choice.

Figure 2.6 Results of variance component analysis using final selected models for depression, cognitive and personality traits in GS20K.



X-axis: names of phenotype; Y-axis: proportion of phenotypic/genetic variance explained by the different components.

I compared the estimates of h_g^2 and h_{gkin}^2 to the publications and found that, for most traits, our results are reasonably close to those in the literature (Table 2.5).

Table 2.5 Comparisons of the results from final selected models in GS20K to previous published results for cognitive, personality and depression traits.

Phenotype	Selected Models		Publications	
	h^2_g (S.E)	h^2_{gkin} (S.E)	h^2_g (S.E)	h^2_{ped} (S.E)
Cognitive				
g	0.23 (0.02)	0.54 (0.03)	20-50% [133,134]	50-80% [135]
Education	0.16 (0.02)	0.44 (0.03)		
Vocabulary	0.26 (0.02)	0.56 (0.03)		
Verbal Fluency	0.19 (0.02)	0.46 (0.03)		
Digit Symbol Test	0.21 (0.02)	0.36 (0.03)		
Logical Memory	0.12 (0.02)	0.32 (0.03)		
Personality				
Neuroticism	0.11 (0.02)	0.30 (0.03)	0-18% [136-138]	34-48% [139]
Extraversion	0.13 (0.02)	0.13 (0.02)		
Depression				
Major Depressive Disorder	0.12 (0.05)	0.47 (0.06)	21-32% [140,141]	~37% [142]

More recently, one colleague from my group (Carmen Amador) further developed this method to study the regional differences in health-related phenotypes and proved that the causes of regional variation are socioeconomics and lifestyle rather than genetics [132] (Publication S2.3).

To conclude, this study reveals new insight into trait components of human complex traits compared to the traditional variance component analysis which mainly focuses on the additive genetics. With appropriate data (a large cohort with multiple degrees of relatives), my method is able to disentangle the confounding genetic and environmental effects shared by relatives, which points out the models to be used in further studies.

Chapter 3: New GWAS Method Correcting for Family Structure

3.1 Introduction

Previously in Chapter 2, I conducted variance component analyses and identified that couple environment, sibling environment and pedigree-associated genetics are three additional major components of human complex trait variations in addition to SNP-associated genetics.

In this chapter, I will apply the discovery of major trait variation contributors into a genome-wide association study (GWAS) by taking account of the corresponding covariance structure contributed by these additional factors in a linear mixed model (LMM), to see whether the performance of this extended method is greater than that of traditional GWAS method which only models SNP-associated genetics.

This work was done in collaboration with a colleague (Oriol Canela-Xandri) from another department of the university. My colleague developed this method and implemented it into a genetic analysis tool named DISSECT, [143]; whereas my role was to perform GWAS on 16 anthropometric and cardio-metabolic traits in GS20K using both the traditional and the extended methods and, afterwards, making comparison of GWAS performance.

Currently, we are in the stage of applying for publication (draft attached, see Publication S3.1) and here the main text of this chapter starts, beginning with a literature review about association studies.

In quantitative genetics, association studies could be classified as family-based association studies and population-based association studies.

Transmission disequilibrium tests (TDT) [144] and modified TDT [145,146] are one of the methods usually applied for family-based association studies. TDT requires

known pedigree, normally nuclear families with at least one affected offspring, as it tests whether the transmission of alleles from parents to offspring is in association with disease risk. The first GWAS was a study of such kind, e.g. genome-wide family-based association study based on modified TDT [53].

However, the acknowledged starting point of GWAS is a publication by Wellcome Trust Case Control Consortium (WTCCC) in 2007 [54] because it is the first population-based association study using genome-wide marker data genotyped by high-coverage SNP chip [55]. The samples required for a population-based association study usually are genetically unrelated individuals of homogeneous ethnicity. The reason for having unrelated individuals originally is due to the cost because genotyping the same number of independent samples (unrelated individuals) provides more power in the analysis than genotyping dependent samples (related individuals) [55,80]; whereas the individuals have to be ethnically homogeneous because population structure (or population stratification) could lead to false positive discoveries [147]. Principal components analysis (PCA) is normally used in population-based GWAS for detection and subsequent removal of individuals who are ethnic outliers in the quality control and for correcting population structure in the analysis [148].

With the development of the SNP chip technique, the cost of genotyping a single individual becomes much cheaper. Therefore, having more and more individuals genotyped (large-cohort study) is becoming more common and, inevitably, there will be participants with known or unknown multiple degrees of relationship recruited into these studies. A method to correct for population structure and cryptic/family relationship simultaneously is fitting the first few principal components as fixed effects and polygenic effects (SNP effects) as a random effect (represented by GRM) in a LMM [149-151]. Close relatives such as nuclear family members are usually removed from the analysis to avoid having false positives due to overweighting information contributed by family structure [152], which can greatly reduce the sample size and hence reduces the power.

In Chapter 2, I have demonstrated that my extended variance component analysis method could disentangle the phenotypic variance of a trait into SNP-associated genetics and familial confounding effects (including pedigree-associated genetics and

common environment shared by either by family members or by partners or by siblings) in the presence of relatives. Here, I will aim to identify trait-associated loci by taking account of these familial effects, including related individuals in the analysis. The rationale is that, on one hand, in the presence of relatives, modelling familial effects could remove the false positive associations due to genetics by environment correlation shared between relatives; on the other hand, by considering phenotypic components other than SNP effects, the residuals shrink, leading to smaller standard errors for the estimates of SNP effect size and thus increasing power. By this, it is possible to maximise the sample size of study population, increase the detection power and get rid of potential artificial association due to familial structure.

In this study, I applied this extended method as well as the traditional GWAS method to 16 traits related to anthropometrics and cardio-metabolism in GS20K, compared the performance, in terms of false and true positives, of this method and the traditional method and checked if any novel associations were uncovered.

3.2 Methodology

3.2.1 Data and Matrices

The cohort (GS20K), covariates and phenotypes used for this study are the same as those used and described in Chapter 2: 2.2.7.2, e.g. 20,032 individuals of recent Scottish descent, 520k autosomal common SNPs after quality control and 16 traits related to anthropometrics and cardio-metabolism.

The matrices ($\mathbf{GRM_g}$, $\mathbf{GRM_{kin}}$, $\mathbf{ERM_{Family}}$, $\mathbf{ERM_{Sib}}$ and $\mathbf{ERM_{Couple}}$) used here are exactly those described in Chapter 2: 2.2.7.3.1 and 2.2.7.3.2.

3.2.2 Models

This is a two-step GWAS approach. In the first step, I estimated the parameter $\mathbf{W}$ (representing the covariance of the mixed effects of all random effects fitted in the model) that best describes the phenotypic variation of a trait. In the second step, I

conducted SNP-association test for each SNP using the eigenvectors and eigenvalues of **W**. After the decomposition of **W**, the computational complexity in step 2 becomes $\sim O(n)$, which enables the model to be fitted simultaneously in different parallel nodes thousands or millions of times repeatedly in a practicable elapsed time.

This two-step GWAS method was built in DISSECT, a free software designed for analysing big genomic datasets via clusters [143] used in this study.

3.2.2.1 Step 1: Estimating **W** Matrix

In the first step, the model used was

$$\mathbf{y} = \mathbf{X\beta} + \sum_i \mathbf{r}_i + \boldsymbol{\varepsilon}, \text{with } \mathbf{V} = \sum_i \mathbf{V}_{r_i}\sigma^2_{r_i} + \mathbf{I}\sigma^2_{\varepsilon},$$

where **y** is an *n* (number of individuals) × 1 vector of observed phenotypes and **V** the total phenotypic variance matrix; **β** is an *m* (the total level of covariates) × 1 vector of fixed effects, including sex, age, age^2, sex-by-age interaction, clinic, standardised SIMD and SIMD^2 and the first 20 eigenvectors of $\mathbf{GRM_g}$ (to alleviate problems associated with data structure), and **X** its design matrix with dimension $n \times m$; $\mathbf{r}_i$ is an $n \times 1$ vector of random effect with $\mathbf{r}_i \sim N(0, \mathbf{V}_{r_i}\sigma^2_{r_i})$; and **ε** is an $n \times 1$ vector of residuals with variance σ^2_{ε}.

By defining **w** and **W** as,

$$\mathbf{w} = \sum_i \mathbf{r}_i \text{ and } \mathbf{W} = \frac{\sum_i \mathbf{V}_{r_i}\sigma^2_{r_i}}{\sum_i \sigma^2_{r_i}},$$

The initial equation could be rewritten as,

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{w} + \boldsymbol{\varepsilon}, \text{with } \mathbf{V} = \mathbf{W}\sigma^2_w + \mathbf{I}\sigma^2_{\varepsilon},$$

where **w** is an $n \times 1$ vector of the mixed effects of all random effects $\mathbf{r}_i$, with $\mathbf{w} \sim N(0, \mathbf{W}\sigma^2_w)$, σ^2_w is the total variance explained by this model with $\sigma^2_w = \sum_i \sigma^2_{r_i}$ and the definitions of the remaining parameters are the same as before.

Therefore, the **W** matrix can be obtained by summing the design matrices of all random effects fitted in the model weighted by the variance explained by each random effects according to variance component analysis.

In this study, the random effects $\mathbf{r}_i$ modelled including the SNP-associated genetic effects $\mathbf{g_g}$ with $\mathbf{g_g} \sim N(0, \mathbf{GRM_g}\sigma_g^2)$, the pedigree-associated genetic effects $\mathbf{g_{kin}}$ with $\mathbf{g_{kin}} \sim N(0, \mathbf{GRM_{kin}}\ \sigma_{kin}^2)$, the common environmental effects shared by nuclear family members $\mathbf{e_f}$ with $\mathbf{e_f} \sim N(0, \mathbf{ERM_{Family}}\ \sigma_{ef}^2)$, the common environmental effects shared by sibling $\mathbf{e_s}$ with $\mathbf{e_s} \sim N(0, \mathbf{ERM_{Sib}}\ \sigma_{es}^2)$ and the common environmental effects shared by partners (members of a couple) $\mathbf{e_c}$ with $\mathbf{e_c} \sim N(0, \mathbf{ERM_{Couple}}\ \sigma_{ec}^2)$.

In Chapter 2, I conducted model selection for variance component analysis and identified the important contributors to trait variation and the proportion of phenotypic variance explained by each of them (Table 2.3). From that, I subsequently computed the **W** matrix based on the estimates of σ_g^2, σ_{kin}^2, σ_{ef}^2, σ_{es}^2 and σ_{ec}^2. For example, $\mathbf{W} = \frac{\mathbf{GRM_g}\times 0.2 + \mathbf{GRM_{kin}}\times 0.2 + \mathbf{ERM_{Couple}}\times 0.1}{0.2+0.2+0.1}$ if the selected model for a trait was model 'GKC' and the estimates of h_g^2, h_{kin}^2 and e_c^2 were 20%, 20% and 10% respectively; and $\mathbf{W} = \mathbf{GRM_g}$ if only SNP effects are considered.

3.2.2.2 Step 2: Test SNP Association using **W** Matrix

The model used for testing the trait-association for each SNP in turn was:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{X_{SNP}}\beta_{SNP} + \boldsymbol{w} + \boldsymbol{\varepsilon}\ with\ \mathbf{V} = \mathbf{W}\sigma_w^2 + \mathbf{I}\sigma_\varepsilon^2$$

where β_{SNP} is the effect size of the reference allele of the SNP being tested for association and $\mathbf{X_{SNP}}$ is an n × 1 genotype vector for that SNP (coded as 0, 1, 2 for having 0, 1 and 2 reference alleles) and the definitions of the remaining parameters are the same as before.

Since **W** and σ_w^2 are fixed parameters known from step 1, it is possible to speed up the association analysis by performing eigen-decomposition of **W** and transform the data to a space where the covariance matrix **W** is diagonal using the eigenvectors of **W**.

The eigen-decomposition of **W** is,

$$\mathbf{W} = \mathbf{\Lambda \Sigma \Lambda^{-1}},$$

where $\mathbf{\Lambda}$ and $\mathbf{\Sigma}$ are the eigenvectors and eigenvalues matrices of $\mathbf{W}$, respectively.

Therefore, the map, $(\cdot)$, to transform the data is,

$$\varphi(\cdot) = \mathbf{\Lambda^{-1} \cdot \Lambda}$$

For example, $\varphi(\mathbf{W}) = \mathbf{\Lambda^{-1} W \Lambda} = \mathbf{\Sigma}$.

By performing eigen-decomposition and data transformation, the distribution of $\mathbf{y}$ changed from the origin from, $\mathbf{y} \sim \mathrm{N}(\mathbf{X\beta} + \mathbf{X_{SNP}}\beta_{SNP}, \mathbf{W}\sigma_w^2 + \mathbf{I}\sigma_\varepsilon^2)$, to that after eigen-decomposition, $\mathbf{y} \sim \mathrm{N}(\mathbf{X\beta} + \mathbf{X_{SNP}}\beta_{SNP}, \mathbf{\Lambda}[\mathbf{\Sigma}\sigma_w^2 + \mathbf{I}\sigma_\varepsilon^2]\mathbf{\Lambda^{-1}})$, and finally to that after data transformation, $\varphi(\mathbf{y}) \sim \mathrm{N}(\varphi(\mathbf{X})\mathbf{\beta} + \varphi(\mathbf{X_{SNP}})\beta_{SNP}, \mathbf{\Sigma}\sigma_w^2 + \mathbf{I}\sigma_\varepsilon^2)$. Since $\mathbf{\Sigma}$ is diagonal, the covariance matrix $(\mathbf{W})$ in the transformed space $(\varphi(\mathbf{W}) = \mathbf{\Sigma})$ is diagonal. Consequently, the computational complexity of fitting the model reduces to $\sim O(n)$, compared to $\sim O(n^3)$ if solving the model in a standard way.

3.2.2.3 Three Different GWAS Methods

In my study, I compared the performance of three different GWAS methods, abbreviated as TU, TR and SR.

In the TU method (Traditional method with Unrelated individuals), I only considered one random effects (SNP-associated genetic effect, represented by $\mathbf{GRM_g}$) in the model and only included 7,370 unrelated individuals (r<0.025, including couples who are not genetically related with one another) in the analysis. This is the GWAS method and type of population frequently used in publications.

In the TR method (Traditional method with Related individuals), I only considered one random effects (SNP-associated genetic effects, represented by $\mathbf{GRM_g}$) in the model but included all ~20k individuals (including related individuals) in the analysis.

In the SR method (extended method based on Selected model per trait with Related individuals), I accounted for all the random effects which have been previously identified as contributors of trait variation for each trait (Table 2.4) in the model with

matrix **W** and used all ~20k individuals in the analysis. Note that the selected model varies depending on traits.

3.2.3 Simulation

This method has been validated by my colleague via simulation study. Details about the simulation study is available in Publication S3.1.

3.3 Results

3.3.1 GWAS Performance Comparison within GS20K

I conducted GWAS in GS20K for 16 anthropometric and cardio-metabolic traits using three different methods, TU (traditional method with unrelated individuals), TR (traditional method with all individuals) and SR (extended method with all individuals).

To compare the GWAS performance diversely for different SNP sets (e.g. discovery power for trait-associated SNPs and false positive rate (FDR) for non-associated SNPs), I classified SNPs according to the p-values obtained from each method and the correlation between SNPs estimated from genotype data.

Since the sample size of my study population is relatively limited (~7k for TU and ~20k for TR and SR) compared to the sample size of published GWAS meta-analyses (~100k), here I defined associations with p-values less than 10^{-5} as suggestive trait-associated SNPs (suggestive SNPs). The remaining associations with p-values larger than 10^{-5} were defined as non-associated SNPs although some of them might be real causal loci undetected due to lack of power.

Subsequently, I filtered the suggestive SNPs detected by each method according to the genotypic correlation between them because no LD pruning was conducted for SNP data prior to GWAS and I wanted to make sure that the suggestive SNPs detected were independent, i.e. causing by different causal variants. For each method, I selected the most significant suggestive SNP (smallest p-value) and then removed any suggestive

SNPs that were correlated with that one ($r^2 \geq 0.7$). Next, I selected the second most significant suggestive SNP that had not been eliminated in the previous process, and repeated the procedure until all suggestive SNPs that remained would no longer be in high correlation with each other ($r^2 < 0.7$) and each of them might represent the signal from different genetic variants.

Furthermore, I checked the genotypic correlation between the suggestive SNPs detected by two different methods to see whether some signals detected by one method were method specific. If a suggestive SNP detected by one method was not highly correlated ($r^2 < 0.7$) with any other suggestive SNPs detected by the other two methods, the signal leading to that suggestive SNP was counted as a unique signal for that method as the underlying genetic variant could not be detected by any other method. On the contrary, if a suggestive SNP detected by one method was highly correlated ($r^2 \geq 0.7$) with one or two suggestive SNPs detected by the other two methods, the signal resulting in those associations was counted as a common signal as the underlying genetic variant could be detected by two or more methods.

3.3.1.1 Comparison of Non-Associated SNPs

After classification, I first compared the p-values for non-associated SNPs obtained from TU, SR and TR methods with each other. I regressed the $-\log_{10}$ p-values of the non-associated SNPs detected in one method against the $-\log_{10}$ p-values for the same SNPs in another method (it is possible that some non-associated SNPs detected in the first method are suggestive SNPs in the second method) for each trait (Table 3.1). As shown in Table 3.1, the regression coefficients for any specified method comparison (TR vs. TU, SR vs. TU and SR vs. TR) are significantly lower than 1 (coefficient estimate + 2 S.E. < 1) for any trait, which suggests that both the extended SR method and the TR method are expected to reduce the false discovery rate (FDR) as, in general, p-values for non-associated SNPs become less significant when increasing the sample size (TR and SR vs. TU). Besides, there is another tiny but statistically significant improvement in FDR when taking account the identified familial effects in the model, in addition to SNP-associated genetic effects (SR vs. TR). Therefore, the extended SR method should have the lowest FDR compared to the other two.

Table 3.1 Method comparison: the estimate of regression coefficient with S.E. by regressing -log10 p-values for non-associated SNPs (p-values < 10^{-5}) detected in one method against -log10 p-values for the same SNPs from another method.

Trait	Method **TR** vs. **TU**			Method **SR** vs. **TU**			Method **SR** vs. **TR**		
	No. Non-Sig. SNPs	Regression Coefficient		No. Non-Sig. SNPs	Regression Coefficient		No. Non-Sig. SNPs	Regression Coefficient	
		Estimate	S.E.		Estimate	S.E.		Estimate	S.E.
Glucose	519,675	0.4111	0.0013	519,675	0.4125	0.0013	519,545	0.9973	0.0001
Waist	519,714	0.4058	0.0013	519,714	0.4039	0.0013	519,622	0.9754	0.0003
Fat	519,712	0.4048	0.0013	519,712	0.4039	0.0013	519,624	0.9807	0.0003
HDL_C	519,676	0.4519	0.0013	519,676	0.4494	0.0013	519,528	0.9826	0.0003
Weight	519,706	0.4167	0.0013	519,706	0.413	0.0013	519,605	0.9757	0.0003
Height	519,683	0.443	0.0013	519,683	0.4141	0.0012	519,492	0.882	0.0004
Hip	519,709	0.4184	0.0013	519,709	0.4172	0.0013	519,616	0.9813	0.0003
DBP	519,721	0.4016	0.0013	519,721	0.4025	0.0013	519,646	0.995	0.0001
WHR	519,724	0.3946	0.0013	519,724	0.3985	0.0013	519,641	0.9879	0.0002
ABSI	519,724	0.3796	0.0013	519,724	0.3837	0.0013	519,644	0.9877	0.0002
HR	519,717	0.4019	0.0013	519,717	0.4047	0.0013	519,628	0.9926	0.0002
BMI	519,710	0.4137	0.0013	519,710	0.4097	0.0013	519,605	0.9764	0.0003
SBP	519,715	0.4146	0.0013	519,715	0.417	0.0013	519,633	0.993	0.0002
TC	519,700	0.4273	0.0013	519,700	0.432	0.0013	519,521	0.9912	0.0002
Creatinine	519,720	0.4342	0.0013	519,720	0.4283	0.0013	519,606	0.9659	0.0004
Urea	519,726	0.4037	0.0013	519,726	0.4048	0.0013	519,625	0.9935	0.0002

3.3.1.2 Comparison of Suggestive Associations

I then evaluated the detection power of each method by counting the number of suggestive SNPs and number of unique and common signals detected per method (Table S3.1). Note, a suggestive SNP is either contributed by a unique signal or a common signal.

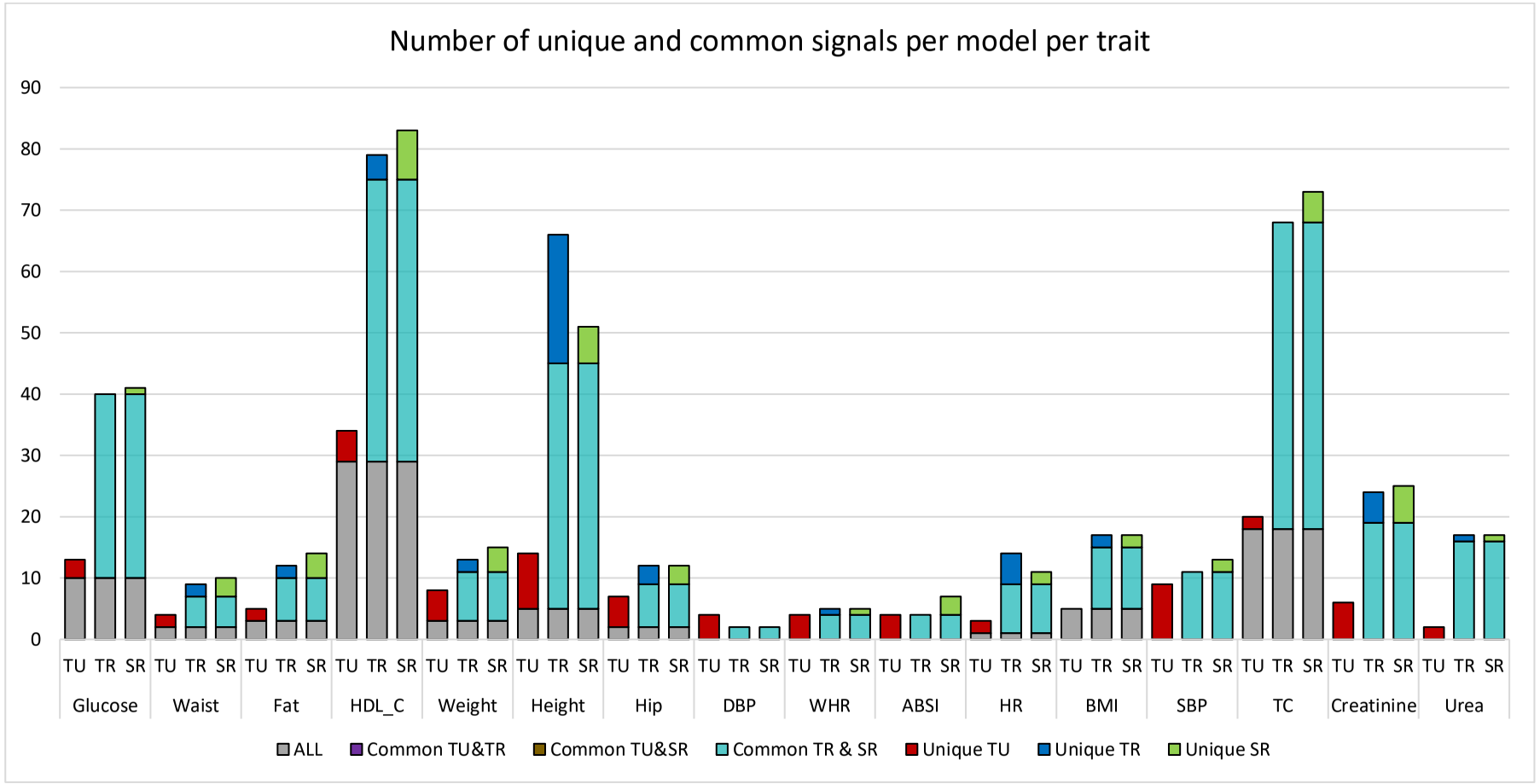
I have summarised the results in Figure 3.1. As shown in Figure 3.1, most suggestive SNPs detected by method TU are due to common signals that can be detected by all methods. Methods TR and SR share another large proportion of common signals due to the inclusion of related individuals in the analyses. The number of unique signals detected by method SR is generally equivalent or higher than method TR, which indicates that detection power is higher if additional familial effects, in addition to SNP-associated genetic effects, are modelled. But a notable discrepancy was found for height, for which the TR method detected 15 more suggestive SNPs than SR method. However, in general, the total number of suggestive SNPs (unique + common signals) detected by each method is TU < TR ≤ SR, i.e. the extended SR method has the overall best detection power.

I took a further look at the detection power of methods TR and SR by comparing the p-values of common signals shared by them. I regressed the $-\log_{10}$ p-values of the common signals obtained from method SR against the $-\log_{10}$ p-values of the same common signals obtained from method TR.

Note, since the suggestive SNP contributed to a common signal in method TR could be different from the suggestive SNP contributed to the same common signal in method SR (but these two different SNPs have to be in high correlation $r^2 \geq 0.7$), here the p-value of a common signal from a method refers to the p-value of the suggestive SNP contributed to that common signal from that method.

As shown in Table 3.2, the regression coefficients for glucose, waist, HDL and BMI are significantly larger than 1 (coefficient estimate - 2 S.E. > 1), which suggests that there is significant evidence for glucose, waist, HDL and BMI that the common signals have lower p-values (more significant) in method SR compared to method TR. The significant opposite trend found for SBP and urea.

Figure 3.1 Number of unique and common hits detected per method per trait



In total, the regression coefficients for 9 out of 15 comparable traits are larger than 1 (although some of them are not significant), which suggests that the detection power of method SR is higher in general.

Table 3.2 Method comparison: regressing -log10 p-values obtained from the SR model on those from the TR model for common signals shared by methods TR and SR

Trait	Number of Common Signals	Regression Coefficient	
		Estimate	S.E.
Glucose	40	1.0088	0.0025
Waist	7	1.1441	0.0489
Fat	9	1.0346	0.0503
HDL	73	1.0152	0.0039
Weight	11	1.0539	0.0411
Height	41	0.9718	0.0629
Hip	9	1.0511	0.0329
DBP	2		
WHR	4	0.6895	0.5823
ABSI	4	0.9859	0.5597
HR	9	1.0136	0.0576
BMI	15	1.0608	0.0215
SBP	11	0.8304	0.0719
TC	68	0.9937	0.0057
Creatinine	19	1.0828	0.1439
Urea	16	0.9174	0.0333

3.3.2 Comparison with Published GWAS Hits

I further evaluated the performance of each method by checking how well my findings (suggestive SNPs) overlap with published GWAS results because the more evidence that can be found to support the trait associations detected by a GWAS method, the more reliable that method seems to be.

Hence, I downloaded a list of published genome-wide (GW) significant hits (p-values $< 5\times10^{-8}$) from the GWAS Catalog (https://www.ebi.ac.uk/gwas/, accessed in Jan-2017 [153]) for each trait (except for ABSI which has no association in the catalogue) and checked the LD (D') between the suggestive SNPs and published GW hits on a reference website http://archive.broadinstitute.org/mpg/snap/ldsearchpw.php. D' is pre-calculated based on the 1000 Genome Pilot 1 database by Johnson et al. [154] and the website only provides LD for SNPs within ± 250kb.

I classified the suggestive SNPs detected in this study based on their locations and LD compared to published GW hits, as well as the p-values obtained from this study. If a suggestive SNP detected in this study locates within +/-250kb of any published GW hits with D' ≥ 0.8, it very likely is a true positive because strong supporting evidence from publications suggests that it is a replication of the published GW hit; For a suggestive SNP detected in our study locating within +/-250kb of any published GW hits with D' < 0.8, there is some evidence from publications suggesting that it potentially is a true positive, e.g. it perhaps is a multiple variant in the same region of the published GW hit; If a suggestive SNP detected in this study locates 250kb away from any published GW hits but it reached GW significance level ($P < 5\times10^{-8}$) itself in any models, it potentially is a true positive because there is some statistical evidence from this study to support that it might be a novel finding; Regarding the remaining suggestive SNPs, there is no evidence from either publications or this study to support that they are true positives.

3.3.2.1 Comparing Levels of Evidence each Method Holds

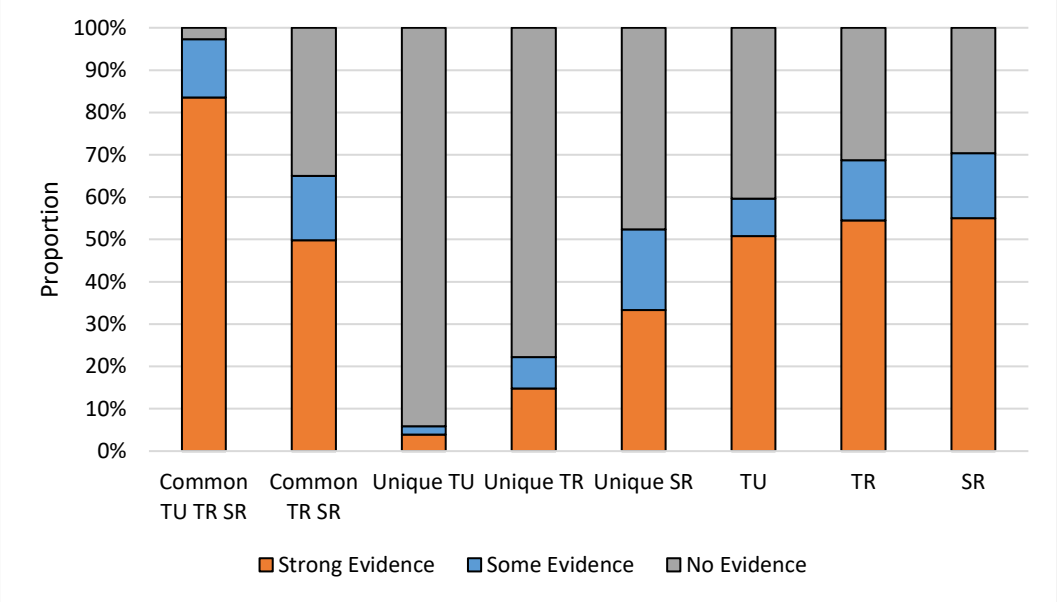
Subsequently, I counted the number of suggestive SNPs detected per method per trait with strong evidence, some evidence and no evidence (Table S3.2) and compared the

overall GWAS performance across traits (excluding height and ABSI) based on the proportion of detected SNPs with and without supporting evidence (Figure 3.2).

According to Figure 3.2, the more methods that detect a common signal, the more likely it is that the association is real. Only ~2.7% of common signals that can be detected by all three methods lack supporting evidence from the GWAS Catalog and this study; whereas the proportion of detected SNPs without supporting evidence increases to 35.0% if the common signals can only be detected by methods TR and SR.

Regarding the unique signals detected by each method (Figure 3.2 and Table S3.2), more than half of (22 out of 42) the unique signals detected by the SR method have strong or some supporting evidence; whereas less than a quarter of (6 out of 27) the unique signals detected by TR method and less than 6% of (3 out of 51) those detected by TU method have supporting evidence.

Figure 3.2 The proportion of suggestive SNPs having strong, some and no supporting evidence per method (and the overlap between methods) across traits, height and ABSI excluded



X-axis: classification of independent suggestive SNPs based on whether the signals contributed were method-specific. Y-axis: the proportion of SNPs that have strong, some and no supporting evidence.

Moreover, by summing unique signals and common signals together, I calculated the overall proportion of detected suggestive SNPs without supporting evidence across traits (excluding height and ABSI) for methods SR, TR and TU, which are 29.6%, 31.43 and 40.3%, respectively (Figure 3.2). In general, the SR method slightly outperforms the TR and TU methods.

However, the method performance varies across traits. Height is the most obvious example where method TR performs significantly better than method SR as method TR could detect 13 more suggestive SNPs with strong supporting evidence from publications (Table S3.2).

3.3.2.2 Comparison of GW Hits Genotyped in GS20K

As I did in Table 3.2 for comparing the p-values of common signals from methods TR and SR, here I compare the p-values of published GW hits from these methods. The rational is that, if the p-values of published GW hits are generally smaller in one of the methods, it is expected to be easier to replicate published findings using that method with an increased sample size.

I extracted SNPs which are on the list of published GW hits downloaded from GWAS Catalog (https://www.ebi.ac.uk/gwas/, accessed in Jan-2017) that happened to be genotyped in GS20K and, subsequently, regressed the $-\log_{10}$ p-values of those SNPs from method SR against the $-\log_{10}$ p-values of the same SNPs from method TR (Table 3.3).

According to Table 3.3, there is significant evidence for glucose, waist, HDL, hip circumference and BMI that published GW hits genotyped in GS20K have lower p-values (more significant) in method SR than method TR because their regression coefficients are significantly larger than 1 (coefficient estimate - 2 S.E. > 1); the significant opposite trend found for height and creatinine.

In total, the regression coefficients for 10 out of 14 comparable traits are larger than 1 (although some of them are not significant), which suggests that the detection power of method SR is higher in general for published GW hits.

Table 3.3 Method comparison: regressing -log10 p-values obtained from the SR model on those from the TR model for genotyped GW hits

Trait	Number of published GW hits Genotyped	Regression Coefficient	
		Estimate	S.E.
Glucose	37	1.0066	0.0012
Waist	34	1.0561	0.0105
Fat	28	1.0153	0.0080
HDL	72	1.0156	0.0029
Weight	12	1.0361	0.0214
Height	187	0.9424	0.0093
Hip	41	1.0300	0.0139
DBP	13	1.0139	0.0213
WHR	20	0.9511	0.0255
ABSI	0		
HR	29	1.0033	0.0069
BMI	116	1.0430	0.0049
SBP	13	0.9989	0.0151
TC	53	1.0094	0.0055
Creatinine	24	0.9052	0.0440
Urea	1		

3.3.2.3 Potential Novel Findings Detected in GS20K

In this study, I detected three potential novel GW hits which reached GW significant level in this study and are located more than 250kb away from any reported GW hits on GWAS Catalog (https://www.ebi.ac.uk/gwas/) when I accessed the website in Jan-2017 (Table 3.4).

The first potential novel finding is for hip circumference and the leading SNP is rs476828 (p-value = 1.22×10^{-8} from SR method) which is located on chromosome 18 and is 2.99Mb away from the closest hip circumference associated SNP (rs12454712) reported from GWAS Catalog. SNPs (rs17782313 and its perfect surrogates including

rs476828) locate ~190kb downstream of melanocortin 4 receptor gene (*MC4R*) known to be strongly associated with childhood (early-onset) obesity [155,156] but evidence for association with hip circumference (or WHR) reported by previous studies is weak [157-159].

The second potential novel discovery is for glucose and the leading SNP is rs7105586 (p-value = 2.58×10^{-8} from SR method) which is located on chromosome 11 and is more than 21Mb away from the closest GWAS hit known to be associated with glucose from the GWAS Catalog. rs7105586 lies ~82kb upstream of leucine zipper protein 2 gene (*LUZP2*). Previous studies have shown that this gene is deleted in some of the Wilms tumour-Aniridia-Genitourinary anomalies-mental Retardation syndrome (WAGR) patients [160] and a fraction of WAGR patients show childhood obesity (which might be related to glucose level in blood) [161,162]. But no evidence for association with glucose is found.

The last potential novel hit is for urea and the leading SNP is rs10480299 (p-value = 1.54×10^{-9} from TR method) which is located inside protein kinase AMP-activated non-catalytic subunit gamma 2 gene (*PRKAG2*) on chromosome 7. This gene is known to be associated with Haemoglobin B level [163] and chronic kidney disease [164].

Recently, a GWAS based on GS20K data, the same data as mine, has been published and, in that study, the association between *PRKAG2* and urea level has been confirmed [165]. However, the other two potential novel hits detected in my study had not been replicated in that one. For the hip circumference hit, that is because the exact trait was not included in that study; whereas for the glucose hit, that probably is because the difference in covariates. In their study, the signal of *LUZP2* (known to be associated with childhood obesity) probably had been removed because they fitted BMI as a covariate in the model.

Table 3.4 Potential novel GWAS findings in our study; their locations, effect sizes (S.E.), leading SNPs, p-values, closest genes and known associations.

Trait	Method	Leading SNP	CHR	Position (bp)	ALLELE	BETA	SE	PV	Gene	Known Association
Hip	**SR**	rs476828	18	60,185,354	A	-0.0043	0.0008	1.22E-08	MC4R	Childhood Obesity [155,156]
	TR	rs476828	18	60,185,354	A	-0.0042	0.0008	5.55E-08		
	TU	rs12964203[a]	18	60,236,371	A	-0.0054	0.0011	1.54E-06		
Glucose	**SR**	rs7105586	11	24,412,117	G	-0.0046	0.0008	2.58E-08	LUZP2	WAGR Syndrome [160]
	TR	rs7105586	11	24,412,117	G	-0.0046	0.0008	3.22E-08		
	TU	null[b]								
Urea	**SR**	rs10480299	7	151,405,818	A	-0.0108	-0.01713	2.70E-09	PRKAG2	Chronic Kidney Disease [164] & Hb Level [163]
	TR	rs10480299	7	151,405,818	A	-0.0109	-0.01735	1.54E-09		
	TU	null[b]								

a. This is a common signal shared by all three methods, but the associated SNP is different in method TU.

b. This is a common signal shared only by methods TR and SR

3.4 Conclusion and Discussion

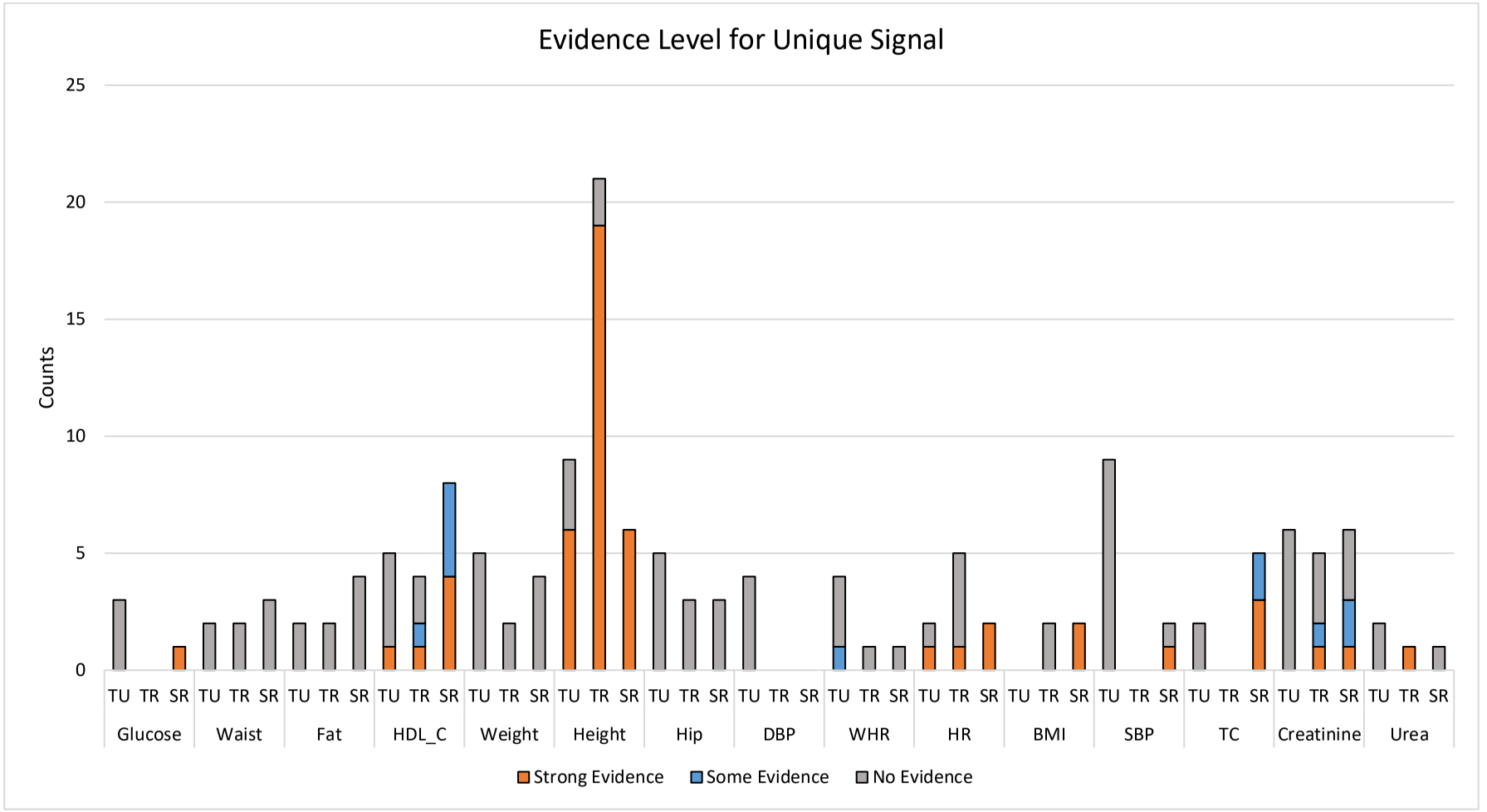
We developed a GWAS method which could take into account multiple random effects. Such a method allows us to include all individuals in an analysis whilst at the same time removing the confounding factors shared among relatives which potentially bias the GWAS results due to overweighting family structure.

I applied this extended GWAS method (SR) on 16 traits related to anthropometrics and cardio-metabolism in GS20K by taking account of all previously identified trait variation contributors as random effects in the GWAS model per trait. The contributors modelled including SNP-associated genetic effects, pedigree-associated genetic effects, common nuclear family environment, common sibling environment and common couple environment. Subsequently, I conducted GWAS on these traits using the traditional methods (TU and TR) which only model SNP-associated genetic effects and compared the GWAS performance.

I observed that the p-values for non-associated SNPs were significantly higher in this extended GWAS method (SR) than the other methods (TU and TR) for any traits (Table 3.1), which indicates that this extended GWAS method is expected to reduce FDR.

The suggestive associations identified in this study were separated into unique signals, which could be detected only by one method, and common signals, which could be detected by more than one methods. Based on the number of unique signals detected by each method and the supporting evidence from GWAS Catalog (https://www.ebi.ac.uk/gwas/), the number of (potential) true positives exclusively detected by this extended GWAS method (SR) is equivalent or higher than those exclusively detected by others (TU and TR) for 12 out of 15 comparable traits (Figure 3.3). In addition, I observed that the overall p-values for common signals were lower in this extended GWAS method than the others for 9 out of 15 comparable traits (Table 3.2); and likewise, the overall p-values for published GW hits genotyped in GS20K were also lower in SR method than the others for 10 out of 14 comparable traits (Table 3.3). These results indicate that this extended GWAS method has a higher detection power in general, compared to the traditional ones.

Figure 3.3 Evidence level for unique signals detected by each method for each trait



Based on a comprehensive method comparison, I demonstrated evidence that this extended GWAS method (SR) which models additional familial effects in addition to SNP-associated genetics has the best overall performance due to lower FDR and higher detection power. This further supports my previous conclusion from Chapter 2 that pedigree-associated genetic effects, couple environment and sibling environment, in addition to SNP-associated genetic effects, are major contributors to human complex trait variance; otherwise, fitting them as random effects in the SR method should not improve the GWAS performance.

However, the performance varies between traits. The most obvious examples being HDL and TC which benefit from modelling extra random effects and height which suffers from modelling extra random effects. Switching from method TR to method SR, 6 and 5 (potential) true associations were detected for HDL and TC respectively, as well as losing 2 (potential) false associations for HDL, with little extra computational cost. However, taking into account of familial effects in GWAS would cost 13 (potential) true findings for height (Figure 3.3).

For traits like height, there is assortative mating [91,166]. Assortative mating generates a positive correlation between the genetic values of partners [93]. Therefore, the genetic similarity between partners is expected to be higher than two random individuals due to assortative mating. In the SR method, the GWAS model used for height was the model 'GKFC', which contained couple resemblance represented by $\mathbf{ERM_{Couple}}$ which had been inaccurately modelled as environment in Chapter 2. The $\mathbf{ERM_{Couple}}$ probably has removed the extra genetic signals contributed by assortative mating in the SR method and thus lead to lower detection power compared to the TR method which did not (Figure 3.3).

To conclude, by applying the knowledge of trait architecture in the GWAS (i.e. fitting familial genetic and environmental effects as well as SNP-associated genetic effects identified to have contributions to trait variations as random effects in a GWAS model), the detection power increased and false-discovery rate decreased, slightly but significantly for some traits. This points out that it is important to study the architecture of human complex traits and, afterwards, using this information for GWAS because it improves GWAS performance with little extra computational cost.

Chapter 4: Predicting Phenotypic Values using Genotype and Genealogy Information

4.1 Introduction

In Chapter 2, I discovered that pedigree-associated genetics, shared couple environment and shared sibling environment, in addition to SNP-associated genetic effects, contribute significantly to trait variation for anthropometric and cardio-metabolic traits in GS20K.

In Chapter 3, I conducted GWAS for these traits in GS20K and found that by adding these additional genetic and environmental factors as random effects in GWAS models, GWAS detection power was boosted and false discovery rate decreased.

In this chapter, I am going to include these factors in prediction models to see whether prediction accuracy is similarly improved.

The aim of this study was to predict the phenotypic values of obesity related traits (including height, BMI, HDL etc.) using kernel ridge regression (KRR) method and predictors that cannot be affected by any traits such as sex, age, genealogy information (e.g. who is your sibling and parent) and genetics (SNP data). The reasons of choosing these measurements as predictors and KRR as method were revealed in the following reviews.

4.1.1 Prediction using Omics Data and Clinic Measurements

For around half a century, medical doctors and scientists have been trying to predict the disease risk for a healthy person and the survival time (rate) for a patient using various sorts of information at hand. However, prior to the omics era, information used for prediction was limited to some clinical traits. Taking cardiovascular diseases (CVD) as an example, by 2013, the most common (~66%) set of predictors used in CVD risk-

assessment models were age, smoking, blood pressure and blood cholesterol; over 90% of the models either included sex or were sex-specific (study men and women separately); whereas only <5% of the studies used genetic data [167], i.e. limited number of genomic prediction studies.

The development of omics technology enables us to measure more types of biological parameters in vivo at relatively low cost. In genetic analysis, omics data can be classified as genomics data (DNA) and the expressed genome data (including epigenomics, transcriptomics, proteomics, metabolomics etc.), both types of omics data can be used as predictors and increase the accuracy of prediction models [69,72,168].

Studies show that including clinical phenotypes like BMI and HDL and expressed genome data like methylation and expression levels in the prediction model could greatly increase the short-term prediction accuracy [69,72], which is useful for diagnosis. However, compared to short-term prediction, long-term prediction on risk-assessment is more helpful for disease prevention. But measurements like BMI, HDL, methylation levels and expression levels could change over time, which are unsuitable for long-term prediction.

To make long-term prediction, relatively constant predictors are required. The stable predictors used in my study included sex, genotype and genealogy information such as who is your sibling and biological parent (which, by default, are fixed from birth) as well as age (which is fixed at the age of prediction).

4.1.2 Difference and Similarity: BLUP, Ridge Regression, KRR and MKL

In this section, I stated the reason of choosing kernel ridge regression in my study by reviewing a few prediction methods and their connections. Methods reviewed included the best linear unbiased predictor (BLUP), ridge regression, kernel ridge regression (KRR) and multiple kernel learning (MKL).

In short, BLUP is a statistic method to estimate random effects in a linear mixed model (LMM) framework [169], commonly used in animal breeding for predicting genetic breeding values; KRR is a prediction method based on ridge regression and kernel trick (details see 4.1.2.3) in machine learning; and MKL is a method that could be implemented in KRR which blends multiple random effects into a combined effects. BLUP with only additive effects is a special form of KRR[170]. However, when the relationship between phenotype and genotype is no longer linear, e.g. non-additive genetic effects such as dominance and epistasis also contribute to the genetic values, the prediction accuracy using KRR is higher than BLUP [171,172].

In my study, I assumed that the phenotype was not only contributed by additive genetic effects, but was also attributable to pedigree-associated genetic effects and familial environmental effects including shared family, couple and sibling environment; and my goal was to predict the phenotypic values contributed by these five effects rather than the genetic breeding values alone. Hence, facing the complexity of trait architecture, I chose KRR over BLUP because KRR should perform better when the relationship between phenotype and predictors is complicated. Additionally, BLUP keeps multiple random effects separately in the model and it needs a dozen of iterations for likelihood convergence; whereas in KRR, all random effects are blended into one by MKL method and no iterative process is needed. Therefore, KRR should be computationally faster than BLUP.

In the following, I reviewed the method of BLUP and KRR, as well as ridge regression and MKL, and their relevance.

4.1.2.1 BLUP

In the simplest model where phenotypes have been pre-corrected for the covariates and only considering additive genetics, the model is

$$\mathbf{y} = \mathbf{g} + \boldsymbol{\varepsilon} \qquad Eq(1)$$

Where $\mathbf{y}$ is n (number of individuals) $\times$ 1 vector of phenotype, $\mathbf{g}$ is $n \times 1$ vector of genetic values (or polygenic effects in human genetics and genetic breeding values in animal breeding) with $\mathbf{g} \sim N(0, \mathbf{A}\sigma_g^2)$, $\mathbf{A}$ being pedigree-based kinship matrix and σ_g^2

being additive genetic variance, and $\boldsymbol{\varepsilon}$ is $n \times 1$ vector of residuals including environmental effects and other factors unaccounted in the model with $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma_\varepsilon^2)$ and σ_ε^2 being residual variance.

Under this model, the variance-covariance matrix of **y**, **V**, is

$$\mathbf{V} = \mathbf{A}\sigma_g^2 + \mathbf{I}\sigma_\varepsilon^2 \qquad Eq(2)$$

Once obtained the estimates of σ_g^2 and σ_ε^2, it is possible to predict **g** and the BLUP of **g** is

$$BLUP(\hat{\mathbf{g}}) = \left[\mathbf{I} + \mathbf{A}^{-1}\frac{\sigma_\varepsilon^2}{\sigma_g^2}\right]^{-1}\mathbf{y} \qquad Eq(3)$$

The genetic values **g** can be considered as the sum of genetic effects of all causal loci, i.e. $\mathbf{g} = \mathbf{X}\boldsymbol{\beta}$, where **X** is an $n \times m$ (number of causal loci) design matrix (genotype matrix) for causal loci and $\boldsymbol{\beta}$ is an $m \times 1$ vector of the effect sizes of causal loci on the trait **y**. Assuming the effect sizes for all causal loci follow the same distribution with $\boldsymbol{\beta} \sim N(0, \mathbf{I}\sigma_\beta^2)$, i.e. the variance explained by a single causal locus is the same on average and equals σ_β^2, then $\sigma_g^2 = m\sigma_\beta^2$. By replacing **g** with $\mathbf{X}\boldsymbol{\beta}$ and σ_g^2 with $m\sigma_\beta^2$, $Eq(2)$ can be rewritten as

$$\mathbf{V} = \frac{\mathbf{X}\mathbf{X}^T}{m}\sigma_g^2 + \mathbf{I}\sigma_\varepsilon^2 \qquad Eq(4)$$

And the BLUP of **g** is

$$BLUP(\hat{\mathbf{g}}) = \left[\mathbf{I} + \mathbf{G}^{-1}\frac{\sigma_\varepsilon^2}{\sigma_g^2}\right]^{-1}\mathbf{y} \qquad Eq(5)$$

Where $\mathbf{G} = \frac{\mathbf{X}\mathbf{X}^T}{m}$, known as the VanRaden G matrix 2 [50], is the identify-by-state based estimator of the true genetic relationship between individuals based on causal loci. **G** can be replaced by GRM (the genetic relationship between individuals estimated by common SNPs across genome) as Yang et al. [51,52] proved that GRM is an estimate of **G**. Note, BLUP using genomic information (usually SNPs) is also called GBLUP.

4.1.2.2 Ridge Regression

Ridge regression is a method based on least squares that deals with the data multicollinearity problem in multiple regression analysis by shrinking the regression coefficients toward 0, e.g. having multiple SNPs in LD with the same causal locus in genomic prediction study. To have a better understanding of the mechanism of ridge regression, here I introduced the basic concept of distance and loss function in statistics.

In mathematics and statistics, the term 'norm' ($\|\cdot\|$) describes the 'length' of a vector in a vector space. In inner product spaces, the norm is defined as the square root of the inner product ($\langle\cdot,\cdot\rangle$), commonly referring to dot product. For example, for vector $\mathbf{x}$, $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{x}}$.

The norm can also be used to quantify the 'distance' between two vectors, e.g. the 'distance' between vector $\mathbf{x}$ and vector $\mathbf{y}$ is $\|\mathbf{y} - \mathbf{x}\| = \sqrt{(\mathbf{y} - \mathbf{x})^{\mathrm{T}}(\mathbf{y} - \mathbf{x})}$. With the concept of 'distance', it is possible to tell whether two vectors are alike and thus solving linear equations.

For example, there are n data points in a data set with the form of $(\mathbf{x_n}, y_n)$, where $\mathbf{x_n}$ is the n^{th} vector of matrix $\mathbf{X}$ (e.g. genotype matrix) and y_n is the n^{th} element of vector $\mathbf{y}$ (e.g. phenotype vector), and the aim is to find the best 'linear regression line' (called 'hyperplane' in machine learning) that describes the most likely linear relationship between data $\mathbf{x_n}$ and y_n, which is by minimising the squared 'distance' between all data points to the hyperplane, $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$.

The equation in the previous example that needs to be minimised, $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$, is called a loss function ($\ell(\cdot)$). However, in studies with big data, very often the solution of $\boldsymbol{\beta}$ is not unique. A regularisation term is added in the loss function to improve the condition of the ill-posed problem by enforcing smoothness [173]; otherwise, the equation system is over-fitted. The final loss function which assigns $\|\boldsymbol{\beta}\|^2$ as a penalty is,

$$\ell(\boldsymbol{\beta}|\lambda) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2 \qquad Eq(6)$$

λ is known as the penalty coefficient and it is a shrinkage parameter. When λ is 0, the solutions are identical to least squares solutions; whereas when λ is infinite, all coefficient estimates in $\boldsymbol{\beta}$ equal 0.

An explicit solution is obtainable by differentiating the loss function, $Eq(6)$, with respect to the unknown vector $\boldsymbol{\beta}$ and setting its derivative to 0. The optimised $\boldsymbol{\beta}$ obtained by this way is called optimiser and this method is known as ridge regression or Tikhonov regularization [174].

Directing to genetics, the loss function of $Eq(1)$ is,

$$\ell(\mathbf{g}|\lambda) = \|\mathbf{y} - \mathbf{g}\|^2 + \lambda\|\mathbf{g}\|^2 \qquad Eq(7)$$

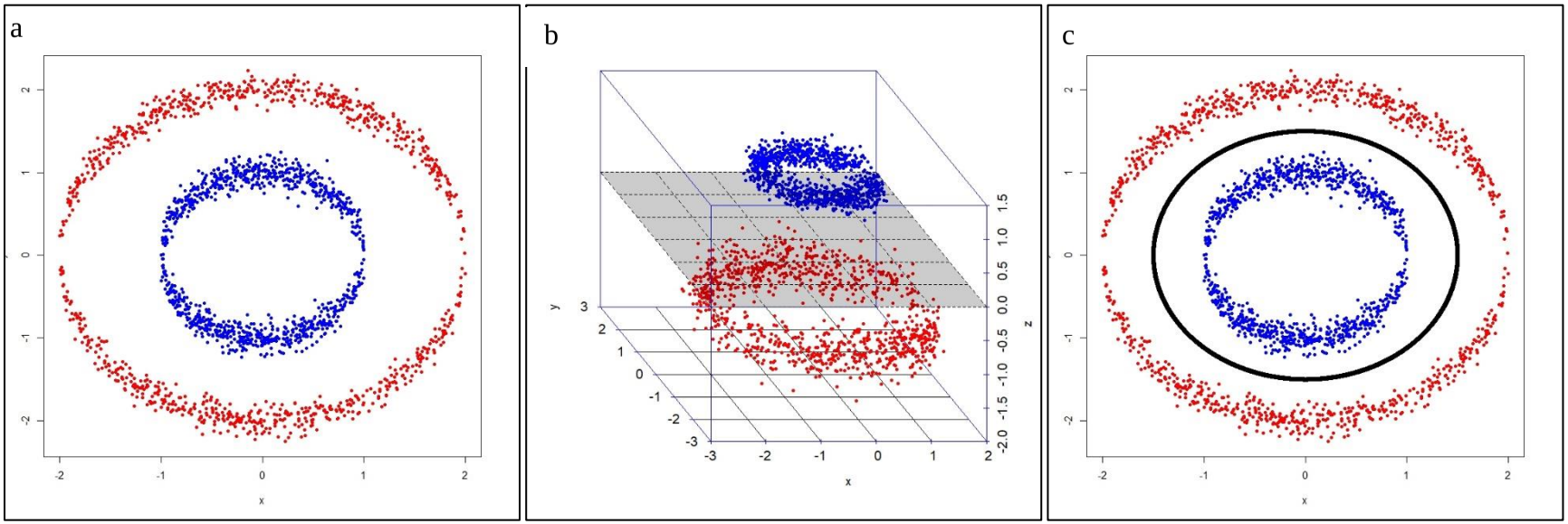
4.1.2.3 Kernel Ridge Regression (KRR)

KRR is a special kind of ridge regression dealing with a non-linear relationship between the data points by using the kernel trick. To understand KRR, I briefly introduced the kernel trick and RKHS theory.

Ridge regression, as any other regression analyses do, reveals the linearity of the data (linear relationship between input and output variables) and traditionally cannot reveal the true relationship between the input and output variables if the relationship between them is non-linear. However, there might be linearity of the data after a certain data transformation.

One typical example is given below. In XY coordinate system, there are 1000 red dots from centred circle with radius of 2 and 1000 blue dots from centred circle with radius of 1. Clearly, in the original 2-D input space (Figure 4.1a), there is no such a regression line that can perfectly separate blue and red dots because their relationship is non-linear. However, if transforming these data points to XYZ coordinate system by setting all blue dots above the XY plane with Z=1 and all red dots below the XY plane with Z=-1, then in the XYZ coordinate system (3-D feature space), the transformed data points are linearly separable and the XY plane (Z=0) is the optimised hyperplane to separate these data points (Figure 4.1b) because it minimises the squared distance between data points and hyperplane.

Figure 4.1 Example of using kernel trick to find hyperplane



Plot a: Red dots are from centred circle with radius of 2 and blue dots are from centred circle with radius of 1 (with some fluctuation).

Plot b: transferring data points into 3-D space by setting z=1 for blue dots and z=-1 for red dots. Hyperplane z=0 in grey.

Plot c: black centred circle with radius of 1.5 is the hyperplane to separate these data points in the original 2-D space.

Transforming the optimised hyperplane obtained in 3-D feature space back to the original 2-D input space, we could get a regression curve (centred circle with radius of 1.5). This regression curve perfectly separates those red and blue dots and clearly reflects how these dots are non-linearly related (Figure 4.1c).

Hence, to deal with a non-linear data set, the feature map ($\Phi(\cdot)$, the function transforms the data points from the input space, where their relationship is non-linear, to the feature space, where the relationship is linear) needs to be known and the inner product (the squared distance) after data transformation needs to be computed. This brings out two problems: first, what the map $\Phi(\cdot)$ is; and second, calculation of the inner product in the high-dimensional feature space takes much more computational time.

RKHS stands for Reproducing Kernel Hilbert Space [175], i.e. the Hilbert Space for a special function and that function is associated with a reproducing kernel. The precise theory of RKHS is beyond this thesis but it helps us to address both issues simultaneously. The idea of using RKHS theory is to help us finding the optimised hyperplane for a set of non-linear data points in its high-dimensional feature space with a kernel function that remains in the original low-dimensional input space. That kernel function can represent the corresponding inner product in high-dimensional feature space, thereby, we neither need to know the map $\Phi(\cdot)$ nor to calculate the inner product in the high-dimensional feature space, e.g. regression curve in Figure 4.1c can be obtained without the data transformation process in Figure 4.1b. This is known as kernel trick in machine learning [176]. Ridge regression using RKHS theory (kernel trick) is called KRR or RKHS regression.

Directing back to quantitative genetics, Gianola et al. first connected RKHS regression with quantitative genetic model [171,177]. Instead of assuming the genetic values $\mathbf{g} = \mathbf{X\beta}$ and restricting that $\mathbf{g}$ and $\boldsymbol{\beta}$ are from normal distribution, they defined the genetic values as a function of genotype, i.e. $\mathbf{g}(\mathbf{X}) = \{g(\mathbf{x_1}), g(\mathbf{x_2}), g(\mathbf{x_3}), \dots, g(\mathbf{x_n})\}^T$, where $\mathbf{x_n}$ is the n^{th} row of genotype matrix $\mathbf{X}$, representing the genotype of the n^{th} individual, and $g(\mathbf{x_n})$ is the genetic value of the n^{th} individual, meaning the average genetic value for infinite individuals having genotype $\mathbf{x_n}$. By replacing $\mathbf{g}$ with $\mathbf{g}(\mathbf{X})$, Eq(1) and its loss function Eq(7) can be rewritten as,

$$\mathbf{y} = \mathbf{g}(\mathbf{X}) + \boldsymbol{\varepsilon} \qquad Eq(8)$$

$$\ell(\mathbf{g}(\mathbf{X})|\lambda) = \|\mathbf{y} - \mathbf{g}(\mathbf{X})\|^2 + \lambda\|\mathbf{g}(\mathbf{X})\|^2 \qquad Eq(9)$$

To find the optimised $\mathbf{g}(\mathbf{X})$ which minimise $Eq(9)$, according to representer theorem from RKHS theory [178], the optimiser $\mathbf{g}(\mathbf{X})$ needs to have the linear form as following:

$$\mathbf{g}(\mathbf{X}) = \sum_{i=1}^{n} \alpha_i \mathbf{K}(\cdot, \mathbf{x}_i) = \mathbf{K}\boldsymbol{\alpha} \qquad Eq(10)$$

$\Phi(\cdot)$ is an unknown feature map that transform data points $\mathbf{X}$ to its unknown RKHS and $\mathbf{K}(\cdot, \mathbf{x}_i) = \langle \Phi(\mathbf{X}), \Phi(\mathbf{x}_j) \rangle$ is the 'distance' between $\mathbf{X}$ and its i^{th} row $\mathbf{x}_i$ in that feature space. $\mathbf{K}$ is an $n \times n$ kernel matrix (or just kernel) with $\mathbf{K}(\cdot, \mathbf{x}_i)$ being its i^{th} row. $\boldsymbol{\alpha}$ is an $n \times 1$ vector of unknown coefficients need to be inferred with α_i being its i$^{\text{th}}$ element. Note, α_i does not have any biological meaning. According to the Moore-Aronszajn theorem (another theorem from RKHS theory), for every positive (semi) definite kernel $\mathbf{K}$ on $\mathbf{X}$, there must be a unique RKHS which makes that $\mathbf{K}$ a reproducing kernel [175]. Therefore, as long as there is a positive (semi) definite kernel $\mathbf{K}$, there is no need to know the map function $\Phi(\cdot)$.

By implementing $\mathbf{g}(\mathbf{X}) = \mathbf{K}\boldsymbol{\alpha}$ and using the property of reproducing kernel that $\|\mathbf{K}\boldsymbol{\alpha}\| = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}\boldsymbol{\alpha}$ in (9) , loss function in $Eq(9)$ could be rewritten as

$$\ell(\boldsymbol{\alpha}|\lambda) = (\mathbf{y} - \mathbf{K}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y} - \mathbf{K}\boldsymbol{\alpha}) + \lambda\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}\boldsymbol{\alpha} \qquad Eq(11)$$

By differentiating the loss function in $Eq(11)$ with respect to the unknown vector $\boldsymbol{\alpha}$ and setting its derivative to 0, an explicit solution is obtained.

$$\hat{\boldsymbol{\alpha}} = (\mathbf{K} - \lambda\mathbf{I})^{-1}\mathbf{y} \qquad Eq(12)$$

Hence, the predicted genetic values are $\mathbf{K}\hat{\boldsymbol{\alpha}}$.

4.1.2.4 Multiple Kernel Learning (MKL)

The MKL method [179] is to create a new kernel by averaging the old ones using the property that the linear sum of kernels is a kernel via the following equation:

$$\mathbf{K} = \frac{1}{\sum_{i=1}^{n} \sigma_{K_i}^2} \sum_{i=1}^{n} \sigma_{K_i}^2 \mathbf{K_i} \qquad Eq(13)$$

Where n is the number of kernels and $\mathbf{K_i}$ is the i^{th} kernel with $\sigma_{K_i}^2$ being its corresponding variance.

By using MKL method, it is possible to blend all the genetic and environmental effects in my model into one kernel because KRR methods do not work with multiple kernels.

4.2 Methodology

This study was a prediction study aiming to predict the phenotypic values, rather than disease risk assessment, for obesity related traits. The study used a 5-fold cross-validation design and ridge regression with a blended kernel computed by genotype and genealogy information (5-fold CV KRR).

I validated the methods by simulation study and then applied this approach on real human data in GS10K (because GS20K data were not ready when I conducted this study in my 1st year).

4.2.1 Data Transformation

In this study, I was interested in predicting the phenotypic values of obesity related traits, including height, BMI, hip circumference, total cholesterol (TC) and HDL-cholesterol level in blood.

I log transformed these traits except for height and, afterwards, adjusted them for sex, age and age^2 by linear regression. Subsequently, I normalised the residuals to standard normal distribution by computing z-score for the residuals.

4.2.2 Kernel Ridge Regression with 5-Fold Cross Validation

I randomly separated GS10K into 5 groups of the same size. Afterwards, I learned the optimiser, $\hat{\boldsymbol{\alpha}}$, using the training data made up of individuals from any 4 out of the 5 groups and made prediction for the validation data consisting of individuals from the 5th group using the following equations.

$$\hat{\boldsymbol{\alpha}} = (\mathbf{K}[Train, Train] - \lambda \mathbf{I})^{-1}\mathbf{y}[Train] \qquad Eq(14)$$

$$\hat{\mathbf{y}}[Validation] = \mathbf{K}[Validation, Train]\hat{\boldsymbol{\alpha}} \qquad Eq(15)$$

$\hat{\boldsymbol{\alpha}}$ is an n_{Train} (number of individuals in the training set) $\times$ 1 vector of unknown coefficients that needs to be solved, $\mathbf{K}$ is an $n \times n$ model-specific kernel matrix, $\mathbf{y}$ is an $n \times n$ vector of normalised phenotype and $\hat{\mathbf{y}}$ is an $n_{Validation}$ (number of individuals in the validation set) $\times$ 1 vector of predicted phenotypic values contributed by genetic and environmental effects included in the model. $Train$ and $Validation$ in $Eq(14)$ and $Eq(15)$ are the index (row and column number) of individuals from training set and validation set in $\mathbf{K}$ and $\mathbf{y}$, respectively.

For each analysis, KRR was repeated five times by setting different groups as training set in each run.

4.2.3 Prediction Models

In addition to sex and age which were adjusted in the first stage of normalising phenotypes, the effects considered in my prediction study in KRR stage were SNP-associated genetic effects, pedigree-associated genetic effects and common environment shared by either nuclear family members, siblings or partners (members of a couple), represented by five design matrices: $\mathbf{GRM_g}$, $\mathbf{GRM_{kin}}$, $\mathbf{ERM_{Family}}$, $\mathbf{ERM_{Sib}}$ and $\mathbf{ERM_{Couple}}$ accordingly.

Based on the combination of effects modelled, there are 31 possible prediction models. As in Chapter 2, I keep using the codes ‘G’ for SNP-associated genetic effects, ‘K’ for pedigree-associated genetic effects, ‘F’ for common family environment, ‘S’ for common sibling environment and ‘C’ for common couple environment, e.g. prediction

model ‘GKC’ is a model that considers SNP- and pedigree- associated genetic effects and common couple environment simultaneously.

Moreover, for each model, I conducted 5-fold CV KRR 8 times with different penalty coefficients λ of 0.001, 0.01, 0.1, 1, 10, 100, 500 and 1000 (to solve the ill-posed problem by smoothing).

5-fold CV KRR for all alternative models, e.g. 31 models × 8 lambdas, was conducted. For results see Table S4.1 for height, Table S4.2 for BMI, Table S4.3 for hip circumference, Table S4.4 for HDL and Table S4.5 for TC.

4.2.4 Estimation of Prediction Accuracy

The prediction accuracy of each model was evaluated based on both the correlation between $\mathbf{y}$ and $\hat{\mathbf{y}}$ and the mean squared error (MSE) $\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$. To distinguish the correlation and the MSE, in this chapter, the results of the former are written in the format of percentage, e.g. 30% rather than 0.3.

Both correlation and MSE give us information about how accurately the prediction model performs; but on different aspects. MSE provides the insight of the deviation between the predicted values and the observed phenotypes; whereas correlation tells how these predicted values and observed phenotypes accord in rank.

Note, in the main manuscript, I excluded λ and only reported the best prediction accuracy (highest correlation and the lowest MSE) obtained for each model, even though λ corresponded to the highest correlation model and the lowest MSE model for the same model may differ.

4.2.5 Kernel Computation

The kernel matrix used for each model was computed by the corresponding design matrices. For details about how design matrices were computed, see Chapter 2.

For a single-effect model, the kernel matrix was the design matrix itself, e.g. the kernel matrix for prediction model 'G' was $\mathbf{GRM_g}$.

For a multiple-effect model, the kernel matrix was computed based on $Eq(13)$, i.e. summing the design matrices according to the variance explained by each of them in S2.4 Table. For example, in variance component analysis, the proportion of phenotypic variance explained by $\mathbf{GRM_g}$, $\mathbf{GRM_{kin}}$ and $\mathbf{ERM_{Couple}}$ were 14.83%, 24.43% and 7.94% respectively in model 'GKC' for total cholesterol. Therefore, the kernel matrix, $\mathbf{K}$, for prediction model 'GKC' for total cholesterol was computed as $\mathbf{K} = \frac{14.83}{47.2}\mathbf{GRM_g} + \frac{24.43}{47.2}\mathbf{GRM_{kin}} + \frac{7.94}{47.2}\mathbf{ERM_{Couple}}$.

Ideally the weights (variance explained by each component) used for blending kernels should be learned within the training data; whereas the weights I used here were learned using the whole population, which would cause an overfitting problem that the prediction accuracy of validation set would be inflated because some of the information from validation set had already been applied in the training process.

However, this study was conducted a few years ago when performing a variance component analysis on a trait in a cohort with size of ~10K with a complex model might take a few hours using old version of GCTA. Therefore, instead of learning the weights for each training set for each model for each trait, e.g. 31×5×5=775 VCA, I assumed that my data are homogeneous, e.g. the effects are the same for the whole population as for any subpopulations, which seems to be a plausible assumption.

4.2.6 Verification of Kernel

Importantly, to make sure KRR method works, I need to prove that my three environmental covariance matrices ($\mathbf{ERM_{Family}}$, $\mathbf{ERM_{Sib}}$ and $\mathbf{ERM_{Couple}}$) are rightful kernel matrices. That is by proving my ERMs are positive semidefinite Gramian matrices.

A Gramian matrix can be written as the form of the inner product of a matrix, e.g. $\mathbf{X} = \mathbf{Y}\mathbf{Y}^{\mathrm{T}}$. My ERMs could be rewritten into such form and thus should be rightful Gramian matrices. For example, an example cohort has 6 individuals (ID) from 3 families. ID

1, 2 and 6 are from Family 1, ID 3 is from Family 2 and ID 4 and 5 are from Family 3. The corresponding $\mathbf{ERM_{Family}}$ for that cohort (left part) and its 'squared root' matrix (right part) are as follow:

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}^{\mathrm{T}}$$

Such format matches the definition of Gramian matrix. In addition, I checked the positive definiteness for my matrices in R and all my ERMs are positive semidefinite. Thereby, my ERMs should be rightful kernel matrices, i.e. positive semidefinite Gramian matrices.

Since GRM is known to be a rightful kernel matrix [170] and the sum of rightful kernel matrices is a rightful kernel matrix, all kernel matrices used in my 31 prediction models should be rightful kernel matrices.

4.2.7 Simulation Study

A simulation study was conducted to test how well the KRR method could perform for phenotypes contributed by different sources of effects. The phenotypes used were from the simulation study in Chapter 2 (S2.6 Table scenario c). They were simulated based on the real genomic and genealogy information in GS10K. For more details about how the phenotypes were simulated see S2.1 Text.

Three sets of simulated phenotypes were used in this study and 10 replicates for each set. In the first scenario, the simulated phenotypes were only contributed by SNP-associated genetic effects, $h_g^2 = 25\%$. In the second scenario, the simulated phenotypes were contributed by both SNP-associated and pedigree-associated genetic effects, $h_g^2 = 25\%$ and $h_{kin}^2 = 15\%$ respectively. In the third scenario, the simulated phenotypes were contributed by four genetic and environmental effects, $h_g^2 = 25\%$, $h_{kin}^2 = 15\%$, $e_c^2 = 15\%$ and $e_s^2 = 5\%$. The remaining proportion of variance was made up by the residual.

For each scenario, I performed KRR using at most two models: a model only including the SNP-associated genetic effects, e.g. prediction model 'G' for all scenarios; and a model matching the trait architecture, e.g. prediction model 'GK' for the second scenario and prediction mode 'GKSC' for the third scenario. A fixed lambda of 1 was used in the simulation study and the prediction accuracy was measured based on the correlation between simulated and predicted values.

In the simulation study, the design matrices $\mathbf{GRM_g}$ and $\mathbf{GRM_{kin}}$ were computed in two ways. In the first case, only causal SNPs from even chromosomes that contributed to SNP-associated genetic effects were used to calculate $\mathbf{GRM_g}$ and those from odd chromosomes that contributed to pedigree-associated genetic effects were used to compute $\mathbf{GRM_{kin}}$. That is assuming the genetic architecture (locations of causal loci) is fully discovered. In the other case, $\mathbf{GRM_g}$ was computed by all common SNPs from even chromosomes and $\mathbf{GRM_{kin}}$ was computed by setting entries in $\mathbf{GRM_g}$ lower than 0.025 to 0. That is assuming the genetic architecture is unknown and some of the causal loci (those contributing to pedigree-associated genetic effects) are completely untargeted by the SNP array.

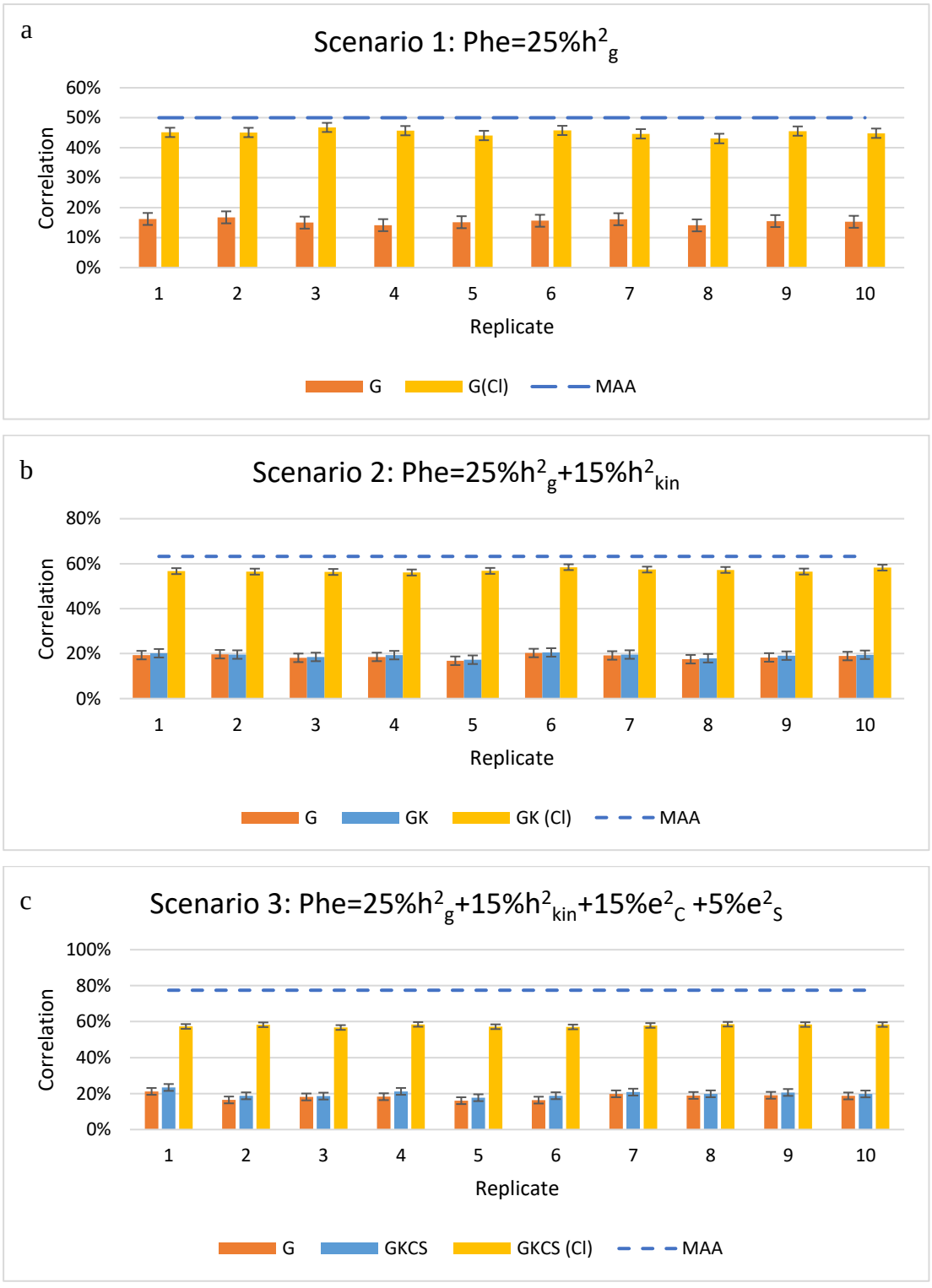
4.3 Results

4.3.1 Simulation

A simulation study was conducted to see how well KRR method could perform under different situations when phenotypes were contributed by different components.

The maximum achievable prediction accuracy for each scenario was $\sqrt{h_g^2} = 50\%$, $\sqrt{h_g^2 + h_{kin}^2} = 63\%$ and $\sqrt{h_g^2 + h_{kin}^2 + e_c^2 + e_s^2} = 77\%$ accordingly; whereas the observed prediction accuracy was ~45%, ~57% and ~58% for the first (Figure 4.2a), the second (Figure 4.2b) and the third scenario (Figure 4.2c) respectively, if the model used in each scenario matched the trait architecture of simulated phenotypes and only causal loci were included in the computation of kernels.

Figure 4.2 Prediction accuracy for simulated phenotypes using 5-fold CV KRR method.



Y-axis: prediction accuracy based on correlation; X-axis: replicate; (Cl): only causal SNPs were contributed to the kernels; MAA: maximum achievable accuracy.

By including non-informative SNPs (non-causal SNPs) into the kernels, the prediction accuracy dropped significantly to ~15%, ~19% and ~20% for the first (Figure 4.2a), the second (Figure 4.2b) and the third scenario (Figure 4.2c) respectively, although models remained unchanged.

Furthermore, there was another ~0.5% and ~1.6% decrease in prediction accuracy for the second (Figure 4.2b) and the third scenario (Figure 4.2c) respectively when effects fitted in the prediction model was no more than SNP-associated genetic effects. The change in prediction accuracy, ~0.5% decrease in Figure 4.2b and ~1.6% decrease Figure 4.2c, due to excluding non-SNP-associated genetic effects in the model was not significant for each replicate, but it was significant if a sign test was conducted for all replicates in each scenario simultaneously (p-value=0.02).

Simulation study thus shows that, the more knowledge we know about the trait architecture (e.g. the effects contribute to the trait variation and the location of causal SNPs), the higher prediction accuracy we could get. However, the proportion of maximal achievable accuracy obtained seems least in scenario 3, which is due to the sparseness of environmental matrices (see 4.4 Conclusion and Discussion)

4.3.2 Prediction for Obesity-Related Traits in GS10K

The simulation study confirms that the accuracy of prediction model benefits from adding extra effects attributable to trait variation in the model, in addition to SNP-associated genetic effects.

Subsequently, I performed 5-fold CV KRR for obesity-related traits in GS10K using a model including either only SNP-associated genetic effects (the base model ‘G’) or all effects previously detected in variance component analysis (VCA) in Chapter 2 that contributed to trait variation (the selected model for each trait in Table 2.2).

On average across traits, the correlation from the selected model was increased by ~1.6% and the MSE from the selected model was reduced by ~0.009, compared to the prediction accuracy for the same trait from the base mode (Table 4.1). However, like simulation study, the improvement in prediction accuracy (i.e. increase in correlation and decrease in MSE) was not significant for each trait due to standard errors.

Table 4.1 Prediction accuracy (S.E.) for base mode and selected VCA model per trait in GS10K.

Trait	Model		Prediction Accuracy	
			Correlation	MSE
Height	Base:	G	37.5% (1.7%)	0.870 (0.013)
	Selected:	GKC	38.7% (1.6%)	0.853 (0.013)
BMI	Base:	G	18.9% (2.0%)	0.964 (0.015)
	Selected:	GKC	21.4% (1.9%)	0.955 (0.014)
Hip	Base:	G	16.1% (2.1%)	0.975 (0.015)
	Selected:	GKC	17.9% (2.0%)	0.963 (0.015)
HDL	Base:	G	21.3% (2.0%)	0.963 (0.015)
	Selected:	GKC	22.8% (2.0%)	0.957 (0.015)
TC	Base:	G	11.7% (2.1%)	0.994 (0.016)
	Selected:	GFS	12.9% (2.1%)	0.992 (0.015)

4.3.3 Prediction Accuracy for Small Groups of Individuals

I further separated GS10K into a few different groups, taking a closer look at how prediction models affected the prediction accuracy for people with different relationships. I extracted the predicted values obtained previously and estimated the prediction accuracy for each group.

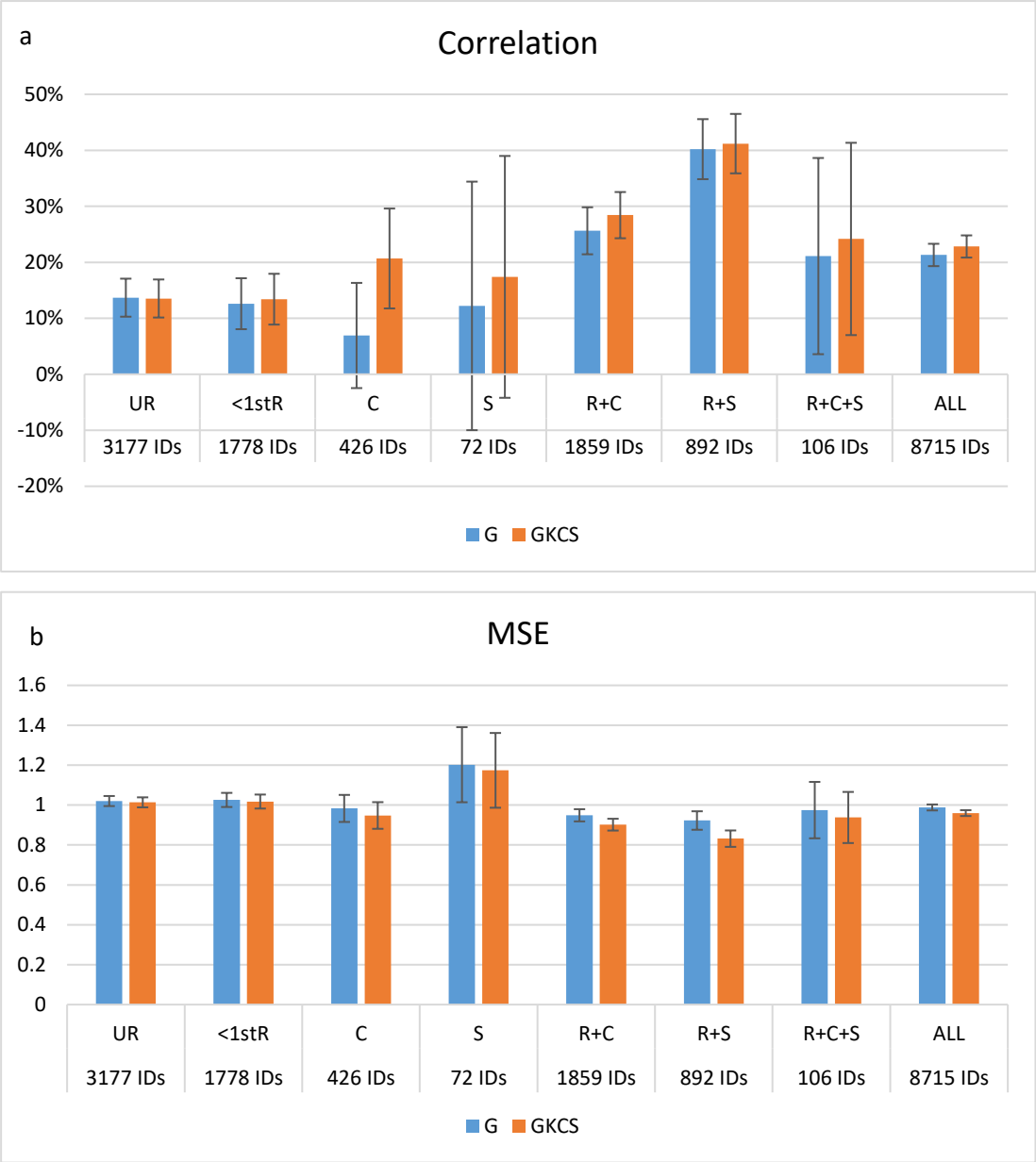
Those groups were: UR (people who are unrelated to anyone else in GS10K), <1stR (people who have at least one non-first-degree relative in GS10K), C (people who have a spouse but no other relatives in GS10K), S (people who have at least one sibling but no other relatives in GS10K), R+C (people who have a spouse and at least one another relative in GS10K), R+S (people who have at least one sibling and one another relative in GS10K) and R+C+S (people who have a spouse, at least one sibling and one another relative in GS10K); the models under comparison were the base model 'G' and the best prediction model 'GKCS' (based on Table S4.4, model 'GKCS' has the highest correlation for HDL); and the example trait taken was HDL.

Figure 4.3a shows the prediction accuracy for different groups of people measured by the correlation. For the same group of individuals (excepting UR group), switching from the base model 'G' to the best model 'GKCS' increased the prediction accuracy, but the amount of increase in prediction accuracy varied upon groups. The maximum increase was 13.75% for group C and the 2nd largest increase was 5.19% for group S; whereas the increase for group <1stR was tiny (0.81%). However, none of the within group difference was significant due to large standard errors. Larger within group differences were observed if prediction accuracy was measured by MSE (Figure 4.3b), e.g. significant difference within group R+S (p-value=0.04).

Regarding the best prediction model 'GKCS', the prediction accuracy of a group of individuals benefited from having more relevant relationships (Figure 4.3a). For example, the prediction accuracy of group R+C and group R+S was significantly higher than that of group <1stR (p-value=5.5×10^{-4} and 4.1×10^{-8} accordingly), the prediction accuracy of group R+C was higher but not significantly than that of group C (p-value=0.35) and the prediction accuracy of group R+S was higher but not significantly than that of group S (p-value=0.25). Similar conclusions hold if prediction accuracies were measured by MSE (Figure 4.3b).

However, the prediction accuracy of group R+C+S was lower than either group R+C or group R+S (Figure 4.3a). That is because the probability of the spouse, sibling and another relative of an individual are all in the training set is very low. Taking the simplest example of having 1 spouse, 1 sibling and 1 another relative in GS10K, the probability that the individual I want to predict is in the validation set while his relatives are all in the training set is $0.8^3\times0.2=0.1024$. Therefore, only 11 individuals (106×0.1024) in group R+C+S benefitted from modelling pedigree-associated genetic effects, common couple environment and common sibling environment simultaneously.

Figure 4.3 Prediction accuracy (S.E.) for different groups of individuals made up of different types of relationship for HDL



Y-axis: prediction accuracy (correlation for plot a and MSE for plot b); X-axis: the name of groups and number of validated individuals (people who have phenotypes) within each group. Definition of each group see manuscript.

4.4 Conclusion and Discussion

In this study, I applied kernel ridge regression (KRR) with a 5-fold cross validation design to predict the phenotypic values for obesity related traits in GS10K. The kernels used were blended kernels created by a multiple kernel learning method based on previous findings of the effects contributing to the trait variation and the amount of phenotypic variance each effect explained. All predictors used were relatively stable measurements, including information on sex, age, genotype and genealogy.

Results indicate that, the prediction ability of obesity-related traits benefits from adding familial effects that were previously identified in variance component analysis (VCA), including pedigree-associated genetic effects, common couple environment, common family and common sibling environment, into the prediction model alongside SNP-associated effects (Table 4.1); whereas the prediction accuracy of individuals benefits from having relatives in the training data who were sharing those familial effects with them (Figure 4.3).

Although the prediction accuracy (referred to correlation between predicted phenotypic values and observed phenotypes, the same below) obtained by my method seems to be relatively low for obesity-related traits (Table 4.1), compared to the high prediction accuracy of ~ 90% for heart disease [180], my results are close to published studies. For published genomic prediction studies including relatives, the prediction accuracy for height, BMI and HDL using whole genome markers are 25-50%, 10-20% and 20-30% respectively [181,182]; whereas in my study the prediction accuracy for height, BMI and HDL from the base model 'G' were 37.5%, 18.9% and 21.3% accordingly.

In genomic prediction, the main problem leading to low prediction accuracy for a trait is lack of knowledge of its causal loci (their locations and effect sizes). Although fitting genome-wide markers could help us to capture the genetic variance contributed by these causal loci due to LD between causal loci and genotyped markers in variance component analysis, adding poor information (genotyped markers that are not causal loci) does not help in terms of estimation of effect size unbiasedly unless a very large discovery set is used [78].

A similar problem was found in my simulation study. A significantly drop of ~30% was observed in prediction accuracy when I predicted the simulated phenotypes (contributed by $h_g^2 = 25\%$) with a kernel computed by both causal SNPs and non-informative SNPs, compared to prediction made upon a kernel computed by only causal SNPs (orange and yellow bars in Figure 4.2a). Another significantly drop of ~38% was detected in prediction accuracy when I predicted the simulated phenotypes (contributed by $h_g^2 = 25\%$ and $h_{kin}^2 = 15\%$) with a kernel computed by non-informative SNPs and some causal SNPs, compared to the result from a kernel computed by all and only causal SNPs (blue and yellow bars in Figure 4.2b).

In fact, the genomic prediction accuracy in the simulation study was reasonable good if the knowledge of genetic architecture of a trait was accurately applied. With a limited sample size of ~10k individuals, ~90% of the maximum achievable prediction accuracy was realised if the kernel used in prediction model contained all causal SNPs and causal SNPs only (yellow bars in Figure 4.2a and Figure 4.2b).

Thereby, the prediction accuracy of obesity-related traits might be further boosted by pre-filtering trait-associated SNPs for computing the GRMs and kernels. For example, a prediction study for HDL shows that KRR with a kernel computed by the top 5 GWAS SNPs sometimes provides higher prediction accuracy than KRR with a kernel computed by all SNPs [181].

Differing from a traditional genomic prediction model, which only considers SNP-associated genetic effects, my extended models also account for the extra similarity between individuals due to pedigree-associated genetic effects and common environment. However, the gain in prediction accuracy by modelling these additional effects was relatively low, e.g. ~1.6% increase for both simulated phenotypes (orange and blue bars in Figure 4.2c) and obesity-related traits in GS10K (Table 4.1). There are three possible explanations for this. First, the main reason is because $\mathbf{GRM_{kin}}$, $\mathbf{ERM_{Sib}}$ and $\mathbf{ERM_{Couple}}$, as well as $\mathbf{ERM_{Family}}$, are very sparse, i.e. have a lot of holes (zero elements) in the matrix. It is difficult to predict with sparse matrix because it is hard to learn the corresponding effects accurately using so few pairs of non-zero elements that are informative; and there will not be a lot of people in the validation set that happen to share such effects with people in the training set that benefit from using

the extended prediction model. Taking couple effects as an example, the non-zero elements (couple pairs) in $\mathbf{ERM_{Couple}}$ is 1,283, which is relatively low compared to ~49M pairs in $\mathbf{GRM_g}$ (Table 2.1). The prediction accuracy of one of the couple in the validation set benefits from modelling couple effects only if the other one of the couple is in the training set. However, separating couples evenly into training and validation sets does not help in terms of learning the couple effects. To learn the couple effects accurately, I need to have as many complete pairs (2 individuals) of couples in the training set as possible, which creates a dilemma. Second, it is hard to learn and predict pedigree-associated genetic effects accurately because the design matrix, $\mathbf{GRM_{kin}}$, was computed by genotyped SNPs which completely excludes any causal variation contributed to this non-SNP-associated genetic effects. Third, in the presence of relatives, a proportion of the additional effects had already been learned and predicted by the base model 'G' due to confounding between the design matrices, e.g. the prediction accuracy of model 'G' increased from ~15.4% for simulated phenotypes only contributed by SNP-associated genetic effects (orange bar in Figure 4.2a) to 18.3% for simulated phenotypes contributed by the same SNP-associated genetic effects and other factors (orange bar in Figure 4.2c). Therefore, the prediction accuracy in base model 'G' was 'inflated' because it was not purely contributed by SNP-associated genetic effects [78].

In this 5-fold CV KRR study, I grouped people at random. Hence, the probability that a couple were both assigned to the training set, one to the training set and the other one to the validation set and both to the validation set are 0.04 (0.2^2), 0.32 ($2\times0.2\times0.8$) and 0.64 (0.8^2) respectively. Hence, the number of individuals that benefitted from modelling couple effects (the second case) was $0.32\times2\times1283=821$, e.g. only ~8% of the whole population. Similarly, the number of individuals that could benefit from modelling pedigree-associated effects and modelling sibling environment were 3172 and 432 respectively (Text S4.1). Therefore, the little improvement, i.e. 1-2% increase in prediction accuracy by adding sparse matrices ($\mathbf{GRM_{kin}}$, $\mathbf{ERM_{Sib}}$ and $\mathbf{ERM_{Couple}}$), may actually be considered as quite good because the increase in the prediction accuracy for the whole population was mainly driven by a small group of individuals (Figure 4.3).

Following from this, I consider that, instead of finding the best prediction model that maximised the general prediction accuracy for the entire population, future studies should focus on applying different prediction strategies and models on different types of people based on what kind of relatives they have in the cohort to maximise the prediction accuracy for each individual or each group because the between group difference in prediction performance for the same model could be huge and should not be neglected. For example, for HDL, the prediction accuracy for people without any relatives in the cohorts was 13.5%, whereas that for people who having sibling and other relationships in the cohorts was over 40% (Figure 4.3). If what I found is correct, then a more reliable prediction accuracy for a person could be easily obtained by recruiting that person's relatives in the study, which might be useful in the aspects of precision medicine.

Chapter 5: Influence of Assortative Mating on Human Complex Traits

5.1 Introduction

Quantitative genetic studies normally follow two key assumptions: first, populations are under random mating; and second, the genetic components of a trait follow an infinitesimal model (that is, they are contributed by an infinite number of causal loci in linkage-equilibrium (LE) across the genome). These assumptions ensure that under this model causal loci are random variables (independent from each other) and there is no covariance structure (also known as population structure or genomic structure in other studies) among them. However, inbreeding, selection, migration, genetic drift, etc., do commonly exist in human history [183-186], all of which could create covariance structure to some extent in real populations and thus potentially violate these assumptions.

Assortative mating is a sexual selection and a non-random mating pattern in which individuals with similar phenotypes mate with each other more frequently than expectation under a random mating pattern [187] and, in this study, the term intensity of assortative mating (ρ) measures how strong such assortment of mate choice is in the population.

Assortative mating generates a positive similarity between members of a couple (partners) and in the absence of other factors, the phenotypic correlation between partners (r_{CP}) is equal to the intensity of assortative mating, i.e. $r_{CP} = \rho$. This mate choice which happens at the phenotype level can be decomposed into genetic and environmental components: ρh^2 of the couple similarity is contributed by the correlation between partners' genetic values, whereas $\rho(1 - h^2)$ of the couple similarity is contributed by the correlation between partners' environmental values [188], unknowingly shared prior to cohabiting after choosing mates. Following

generations of strong assortative mating, the genetic variance of a trait could double compared to its original genetic variance in the founder population [188].

In traditional pedigree studies of heritability, such as MZ-DZ twin studies, populations are assumed to be under random mating [189,190], which might lead to biased results for the trait under study if they are not [191]. Taking height in GS:SFHS as an example, based on my estimation, the point estimate of heritability for height obtained from parent-offspring regression without consideration of assortative mating is 104%, which suggests that it is crucial to know whether the trait is under assortative mating and make necessary adjustment accordingly.

However, to adjust for the influence of assortment of mate choice on the genetic variance of a trait, the intensity of assortative mating needs to be known. For pedigree studies of heritability which consider assortative mating [192,193], it is common practice to replace the intensity of assortative mating (ρ) with the observed phenotypic correlation between partners (r_{CP}). The underlying assumption for this practice is that the couple correlation is purely due to assortative mating, which probably is true for height as one's stature changes little after reaching adulthood; but regarding other obesity related traits like BMI, the couple correlation might be contributed by both mate choice before cohabitation and common couple environment such as exercise and diet after cohabitation [92,194,195]. Therefore, the observed couple correlation should not be taken as the intensity of assortative mating for estimating heritability if the couple similarity is not contributed by assortative mating solely.

A similar problem also potentially affects my variance component study in Chapter 2. In my previous variance component study, I have shown that couple effects are one of the major components of anthropometric and cardio-metabolic trait variations. However, the design matrix for couple effects, $\mathbf{ERM}_{\mathbf{Couple}}$, was a similarity matrix which was originally designed for measuring the similarity between partners due to shared couple environment after cohabiting. This means, if the couple similarity is mainly the result of assortative mating (i.e. $r_{CP} = \rho$), then it no longer reflects pure environmental effects as h^2 proportion of the phenotypic correlation between partners is contributed by genetics. In such a case, adding the couple effect component to the analysis model that already included other genetic and environmental components

leads to little residual variance for traits under assortative mating such as height (Table 2.3).

However, owing to how my data, as well as many other data, were collected (i.e. at a single point and not like a long-term study in which participants were measured for the same phenotypes repeatedly for many years), I am not able to separate assortative mating and couple environmental effects by looking at the changes in observed couple similarity of the phenotype before and after cohabiting. Therefore, except for height for which it is fairly certain that all couple effects shall be contributed by assortative mating [91,166], it is hard to know if there is assortment in mate choice and to what extent the couple similarity is contributed by assortative mating rather than sharing a common "joint lifestyle" environment for other traits.

Previous theoretical studies [188,196] have illustrated the consequence of assortative mating in the field of population genetics, such as how the expectations of genetic variance, heritability and familial resemblance increase under assortative mating at the population level. However, as yet the expected resemblance of in-law relationships other than partners is not revealed, such as the relationship between one's full-sibling and one's spouse (full-sib-in-law relationship, FSIL) and the relationship between one's parent and one's spouse (parent-offspring-in-law relationship, POIL).

In this study, I mathematically derived the expected phenotypic, genetic and environmental resemblance between FSIL and between POIL under assortative mating. I found that the expected phenotypic correlations between FSIL (r_{FSIL}) and between POIL (r_{POIL}) should be positive and are linked with the intensity of assortative mating and heritability.

The aim of this chapter is to test whether there is evidence of assortative mating for the traits investigated and to estimate the heritability and the intensity of assortative mating, by a novel pedigree study using resemblance between in-law relationships derived. Afterwards, I compared the estimate of intensity of assortative mating and the observed phenotypic correlation between partners for the same trait, to see whether there is evidence for other effects contributing to the couple similarity.

This study is supported by simulation studies, with some evidence from real data.

5.2 Assortative Mating Theory

5.2.1 Fundamental Theory

At the start, I review here the fundamental theory relating to assortative mating written in the two standard textbooks for quantitative geneticists, *Genetics and Analysis of Quantitative Traits* and *Introduction to quantitative genetics* [188,196].

According to the theory, if a polygenic trait is contributed by sufficient number of originally unlinked causal loci and individuals mate assortatively for that trait over generations with the same intensity ρ, the genetic and phenotypic variance will increase from $VarG_0$ and $VarP_0$ in the founder population to $VarG_{AM} = \frac{1}{1-\rho h_{AM}^2} VarG_0$ and $VarP_{AM} = \frac{1}{1-\rho h_{AM}^4} VarP_0$ in assortative mating populations at equilibrium, respectively [196]. Note, subscript '0' refers to the founder generation in which the causal loci are in LE; whereas subscript 'AM' represents its subsequent generations that reached equilibrium at assortative mating, i.e. the genetic variance no longer increases.

Hence, the ratio of genetic variance $VarG_{AM}$ to phenotypic variance $VarP_{AM}$ gives a formula that links h_{AM}^2 with h_0^2, which is $h_{AM}^2 = \frac{1-\rho h_{AM}^4}{1-\rho h_{AM}^2} h_0^2$ (Note, the sign in the denominator of equation (9) of Table 10.6 in book [196] should be '-' rather than '+'). By knowing any two out of these three parameters (heritability in the founder population h_0^2, heritability in assortative mating population at equilibrium h_{AM}^2 and intensity of assortative mating ρ), it is possible to derive the third parameter and get a unique solution.

$$h_{AM}^2 = \frac{1 - \sqrt{1 - 4\rho h_0^2 (1 - h_o^2)}}{2\rho(1 - h_o^2)} \qquad Eq1$$

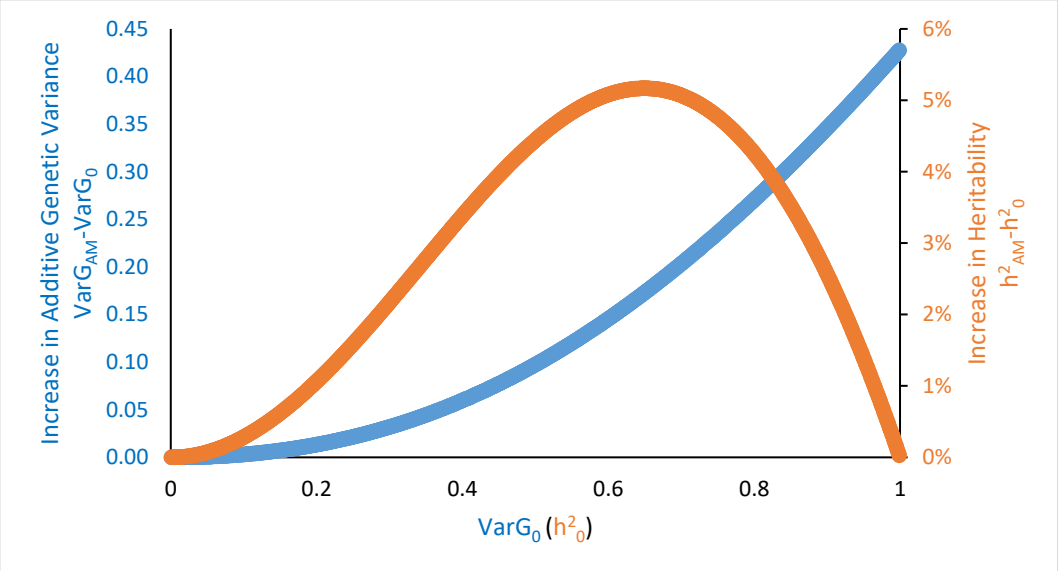
$$h_0^2 = \frac{1 - \rho h_{AM}^2}{1 - \rho h_{AM}^4} h_{AM}^2 \qquad Eq2$$

$$\rho = \frac{h_{AM}^2 - h_0^2}{(1 - h_0^2) h_{AM}^4} \qquad Eq3$$

Based on equations demonstrated above, as well as equations from textbook [196], the key parameters for assortative mating studies are ρ, $VarG_0$ and h_0^2 (or ρ, $VarG_0$ and $VarP_0$) as the increase in additive genetic variance ($VarG_{AM} - VarG_0$) and increase in heritability ($h_{AM}^2 - h_0^2$) vary depending on them.

To explore how these key parameters interact under assortative mating, in Figure 5.1, I plot the increase in additive genetic variance ($VarG_{AM} - VarG_0$) and the increase in heritability ($h_{AM}^2 - h_0^2$) due to assortative mating against a range of $VarG_0$ and h_0^2 for a fixed value ρ of 0.3. Figure 5.1 shows that, for a fixed ρ, the increase in genetic variance ($VarG_{AM} - VarG_0$) increases with $VarG_0$, while the increase in heritability ($h_{AM}^2 - h_0^2$) goes up and then down as h_0^2 increases. The maximum increase in heritability ($h_{AM}^2 - h_0^2$) is 5.17% and that is when the h_0^2 of the trait equals to 64.9%.

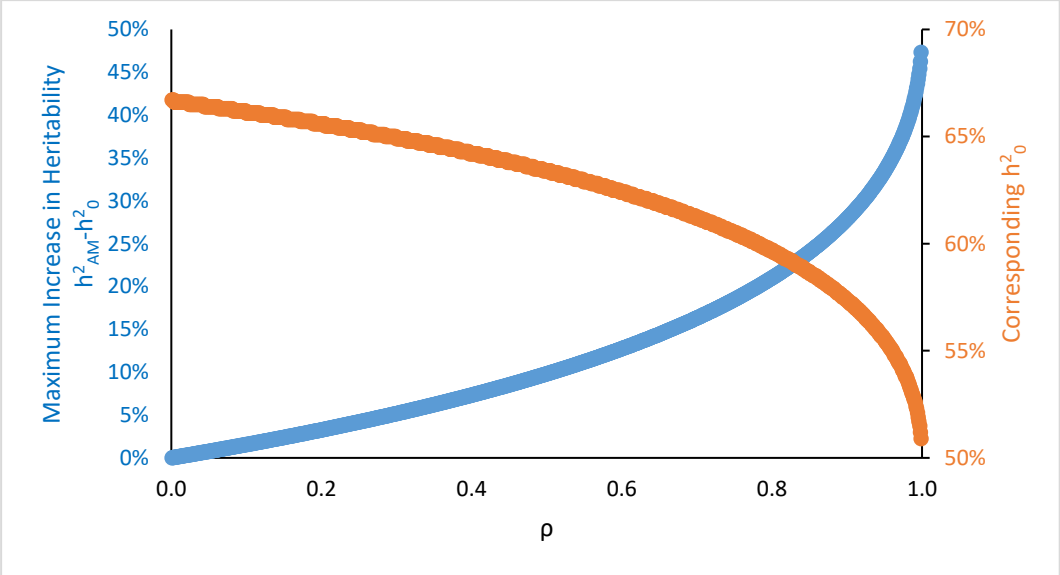
Figure 5.1. The influence of assortative mating on genetic variance and heritability when the intensity of assortative mating equals 0.3.



Y-axis(left): Increase in additive genetic variance due to assortative mating, corresponding to blue curve; Y-axis (right): Increase in heritability due to assortative mating, corresponding to orange curve; X-axis: $VarG_0$ or h_0^2, they are identical here.

In Figure 5.2, I plot the maximum increase in heritability ($h^2_{AM} - h^2_0$) and its corresponding h^2_0 for ρ ranging from 0 to 1. The maximum $h^2_{AM} - h^2_0$ increases and the corresponding h^2_0 declines as ρ increases. The h^2_0 corresponding to the maximum increase in heritability ranges from 66.7% to 50.9%. Hence, traits whose h^2_0 is within that range are influenced by assortative mating to the greatest extent. When ρ is large, the heritability of a trait could be increased by over 40% due to assortative mating. However, ρ for real phenotypes is usually small to moderate and no more than 0.3. Consequently, the maximum increase in heritability is not expected to be more than 5.17% (i.e., relatively small). But, the increase in genetic variance could be large, e.g. the genetic variance increases from 0.6 to 0.75 (25% increase) for a trait whose $VarG_0 = 0.6$ and ρ =0.3 under assortative mating (Figure 5.1).

Figure 5.2. The maximum influence of assortative mating on heritability across different degrees of assortment in mate choice.



Y-axis(left): Maximum increase in heritability for different ρ, corresponding to blue curve; Y-axis (right): h^2_0 corresponds to maximum increase in heritability, corresponding to orange curve; X-axis: intensity of assortative mating ρ.

5.2.2 Increased Resemblance between in-Law Relatives under Assortative Mating

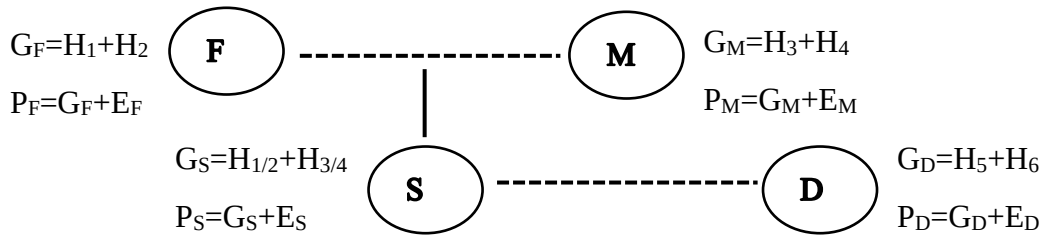
Quantitative genetic textbooks [188,196] not only demonstrate how genetic variance and heritability change under assortative mating but also show how resemblances between different types of blood relatives change accordingly, i.e. how genetic, environmental and phenotypic correlations between blood relatives change under assortative mating compared to the founder population (or any random mating populations). For example, the phenotypic correlation between parents and offspring increases from $\frac{h_O^2}{2}$ in the founder population (or random mating populations) to $\frac{1+\rho}{2}h_{AM}^2$ in populations at equilibrium of assortative mating; and from $\frac{h_O^2}{2}$ to $\frac{1+\rho h_{AM}^2}{2}h_{AM}^2$ for full-siblings accordingly.

Assortative mating also increases the resemblance between in-law relatives, such as partners. However, how resemblance for in-law relatives other than partners changes under assortative mating has been less well studied and I am going to explore these changes here.

I assume that the phenotype of a trait is only contributed by additive effects and the individual environment. Regarding additive effects, it is contributed by an infinite number of causal loci and these loci are physically independent (i.e. recombination rate equals 50%). It is assumed that each individual chooses a mate assortatively with intensity of ρ and furthermore the population has reached equilibrium under assortative mating (genetic variance stops increasing).

5.2.2.1 The Resemblance between Parents and Offspring-in-Law.

First, I derive the parent- and offspring-in-law (POIL) relationship, that is the relationship between one's parents and one's spouse. The example pedigree is as follows:



There are four individuals in this example pedigree, who are the father (F), the mother (M), the son (S) and the daughter-in-law (D). A dotted line means in-law relationship and a solid line means biological relationship. Letters 'P', 'G' and 'E' refer to the phenotypic, genetic and environmental values respectively; whereas letter 'H' refers to the genetic value contributed by causal alleles from one of the individual parental origins. For example, for the father (F), the genetic values contributed by causal alleles from his paternal origin and maternal origin are H_1 and H_2 respectively. Thus, the total genetic value of the father (F) is $G_F = H_1+H_2$. Regarding the son (S), the genetic value contributed by causal alleles transmitted from his father is $H_{1/2}$, which means it is a mixture H_1 and H_2. The difference between G_F and $H_{1/2}$, $G_F - H_{1/2}$, is the father's genetic value contributed by non-transmitted causal alleles. Both transmitted alleles and non-transmitted alleles should contribute equally to the father's total genetic value. The same applies to $H_{3/4}$, the genetic value inherited from the mother.

According to [188,196], the genetic, environmental and genetics-by-environment correlations between partners are ρh_{AM}^2 , $\rho(1-h_{AM}^2)$ and $\rho\sqrt{h_{AM}^2(1-h_{AM}^2)}$ respectively. Since causal alleles from paternal origin and maternal origin shall have equal contribution to one's total genetic effect, therefore the covariance between the father's transmitted genetic value ($H_{1/2}$) and the daughter-in-law's total genetic value (G_D) is:

$$Cov\left(H_{1/2}, G_D\right) = \frac{1}{2}Cov(G_S, G_D) = \frac{1}{2}\rho h_{AM}^2 VarG_{AM} \qquad Eq4$$

And the covariance between the father's transmitted genetic value ($H_{1/2}$) and the daughter-in-law's environmental value (E_D) is:

$$Cov\left(H_{1/2}, E_D\right) = \frac{1}{2}Cov(G_S, E_D) = \frac{1}{2}\rho\sqrt{h_{AM}^2(1-h_{AM}^2)VarG_{AM}VarE_{AM}} \qquad Eq5$$

$Eq4$ and $Eq5$ indicate that, the genetic value contributed by the father's transmitted genome is directly correlated with the daughter-in-law's genetic and environmental values.

Regarding the genetic value contributed by the father's non-transmitted genome, it has a correlation of ρh_{AM}^2 with the genetic value contributed by the mother's transmitted genome due to assortative mating between the father (F) and the mother (M).

$$Cor\left(G_F - H_{1/2}, H_{3/4}\right) = Cor(G_F, G_M) = \rho h_{AM}^2 \qquad Eq6$$

Therefore, the covariance between the father's non-transmitted genetic value ($G_F - H_{1/2}$) and the daughter-in-law's total genetic value (G_D) is:

$$Cov\left(G_F - H_{1/2}, G_D\right) = Cov\left(\rho h_{AM}^2 H_{3/4}, G_D\right) = \rho h_{AM}^2 Cov\left(H_{3/4}, G_D\right)$$

$$= \frac{1}{2}\rho h_{AM}^2 Cov(G_S, G_D) = \frac{1}{2}(\rho h_{AM}^2)^2 Var G_{AM} \qquad Eq7$$

And the covariance between the father's non-transmitted genetic value ($G_F - H_{1/2}$) and the daughter-in-law's environmental value (E_D) is:

$$Cov\left(G_F - H_{1/2}, E_D\right) = Cov\left(\rho h_{AM}^2 H_{3/4}, E_D\right) = \rho h_{AM}^2 Cov\left(H_{3/4}, E_D\right)$$

$$= \frac{1}{2}\rho h_{AM}^2 Cov(G_S, E_D) = \frac{1}{2}\rho^2 h_{AM}^2 \sqrt{h_{AM}^2(1 - h_{AM}^2) Var G_{AM} Var E_{AM}} \qquad Eq8$$

$Eq7$ and $Eq8$ indicate that, the genetic value contributed by father's non-transmitted genome is indirectly correlated with the daughter-in-law's genetic and environmental values through the son's genetic value of maternal origin ($H_{3/4}$). By adding $Eq4$ and $Eq7$ and $Eq5$ and $Eq8$, it is possible to obtain the covariance between the father's total genetic value (G_F) and the daughter-in-law's total genetic value (G_D) and the covariance between the father's total genetic value (G_F) and the daughter-in-law's environmental value (E_D) respectively.

$$Cov\left(H_{1/2}, G_D\right) + Cov\left(G_F - H_{1/2}, G_D\right) = Cov\left(H_{1/2} + G_F - H_{1/2}, G_D\right)$$

$$= Cov(G_F, G_D) = \frac{1}{2}\rho h_{AM}^2 Var G_{AM} + \frac{1}{2}(\rho h_{AM}^2)^2 Var G_{AM}$$

$$= \frac{1+\rho h_{AM}^2}{2} \rho h_{AM}^2 VarG_{AM} \qquad Eq9$$

And

$$Cov(H_{1/2}, E_D) + Cov(G_F - H_{1/2}, E_D) = Cov(H_{1/2} + G_F - H_{1/2}, G_D)$$

$$= Cov(G_F, E_D)$$

$$= \frac{1}{2}\rho\sqrt{h_{AM}^2(1-h_{AM}^2)VarG_{AM}VarE_{AM}} + \frac{1}{2}\rho^2 h_{AM}^2\sqrt{h_{AM}^2(1-h_{AM}^2)VarG_{AM}VarE_{AM}}$$

$$= \frac{1+\rho h_{AM}^2}{2}\rho\sqrt{h_{AM}^2(1-h_{AM}^2)VarG_{AM}VarE_{AM}} \qquad Eq10$$

$Eq9$ and $Eq10$ reveal the genetic and genetics-by-environment correlation between POIL, respectively.

Similarly, although the father's environmental effect (E_F) is not directly correlated with the daughter-in-law's environmental effect (E_D), they are correlated with one another through the son's genetic value of maternal origin ($H_{3/4}$). Thus, the covariance between the father's and the daughter-in-law's environmental values (E_F and E_D) is:

$$Cov(E_F, E_D) = \frac{\sqrt{VarE_{AM}}}{\sqrt{VarG_{AM}}} Cov\left(\rho\sqrt{h_{AM}^2(1-h_{AM}^2)}\, H_{3/4}, E_D\right)$$

$$= \frac{\sqrt{VarE_{AM}}}{\sqrt{VarG_{AM}}}\rho\sqrt{h_{AM}^2(1-h_{AM}^2)}Cov(H_{3/4}, E_D)$$

$$= \frac{1}{2}\frac{\sqrt{VarE_{AM}}}{\sqrt{VarG_{AM}}}\rho\sqrt{h_{AM}^2(1-h_{AM}^2)}Cov(G_S, E_D)$$

$$= \frac{\rho^2 h_{AM}^2(1-h_{AM}^2)}{2} VarE_{AM} \qquad Eq11$$

And the covariance between the father's environmental value (E_F) and the daughter-in-law's total genetic value (G_D) is:

$$Cov(E_F, G_D) = \frac{\sqrt{VarE_{AM}}}{\sqrt{VarG_{AM}}} Cov\left(\rho\sqrt{h_{AM}^2(1-h_{AM}^2)}\, H_{3/4}, G_D\right)$$

$$= \frac{\sqrt{VarE_{AM}}}{\sqrt{VarG_{AM}}}\rho\sqrt{h_{AM}^2(1-h_{AM}^2)}Cov(H_{3/4}, G_D)$$

$$= \frac{1}{2}\frac{\sqrt{VarE_{AM}}}{\sqrt{VarG_{AM}}}\rho\sqrt{h_{AM}^2(1-h_{AM}^2)}Cov(G_S, G_D)$$

$$= \frac{\rho^2 h_{AM}^2}{2}\sqrt{h_{AM}^2(1-h_{AM}^2)VarG_{AM}VarE_{AM}} \qquad Eq12$$

$Eq11$ and $Eq12$ indicate the environmental and environment-by-genetics correlation between POIL, respectively. Therefore, by adding $Eq\,9$, $Eq\,10$, $Eq\,11$ and $Eq\,12$ together, it is possible to get the phenotypic covariance and correlation between POIL, which is:

$$Cov(G_F, G_D) + Cov(G_F, E_D) + Cov(E_F, E_D) + Cov(E_F, G_D)$$

$$= Cov(G_F, G_D + E_D) + Cov(E_F, G_D + E_D)$$

$$= \ Cov(G_F + E_F, G_D + E_D)$$

$$= Cov(P_F, P_D)$$

$$= \frac{1+\rho h_{AM}^2}{2}\rho h_{AM}^2 VarG_{AM} + \frac{1+\rho h_{AM}^2}{2}\rho\sqrt{h_{AM}^2(1-h_{AM}^2)VarG_{AM}VarE_{AM}}$$

$$+\frac{\rho^2 h_{AM}^2(1-h_{AM}^2)}{2}VarE_{AM} + \frac{\rho^2 h_{AM}^2}{2}\sqrt{h_{AM}^2(1-h_{AM}^2)VarG_{AM}VarE_{AM}}$$

$$= \frac{1+\rho}{2}\rho h_{AM}^2 VarP_{AM} \qquad Eq13$$

And

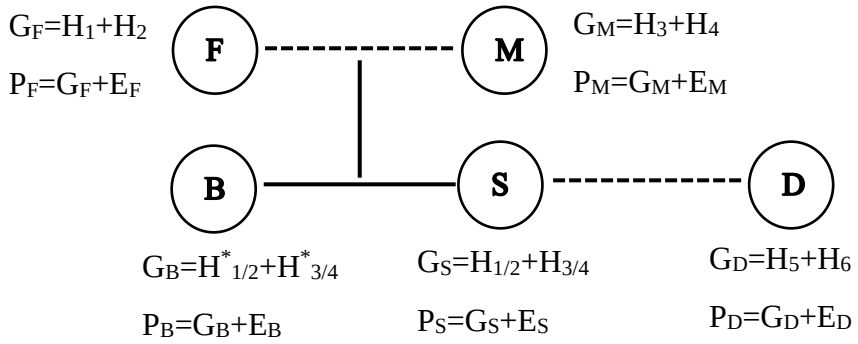
$$Cor(P_F, P_D) = \frac{1+\rho}{2}\rho h_{AM}^2 \qquad Eq14$$

Thus, under assortative mating, the expected phenotypic correlation between POIL is $\frac{1+\rho}{2}\rho h_{AM}^2$, which is the product of the expected phenotypic correlation between parent-offspring ($\frac{1+\rho}{2}h_{AM}^2$) and the expected phenotypic correlation between partners (ρ).

With ρ of 0.3 and h^2_{AM} of 80%, the phenotypic correlation between POIL is 0.156, much higher than the expected correlation of 0 under random mating.

5.2.2.2 The Resemblance between Sib and Sib-in-Law

The second example refers to sib- and sib-in-law (FSIL) relationship, which is the relationship between one's full-sibling(s) and one's spouse. The example pedigree is as follows.



A brother (B) is added into the previous example pedigree. The terminology is the same as before. Since full-siblings share 50% of the genome, I further separate their total genetic values into shared part and non-shared parts. For example, $H_{1/2}$= $H_{1/2(shared)}$+ $H_{1/2(unique)}$ and $H^*_{1/2}$= $H_{1/2(shared)}$+ $H^*_{1/2(unique)}$. Both shared and non-shared parts should contribute equally to one's total genetic value. The same rules apply for $H_{3/4}$ and $H^*_{3/4}$, the genetic value inherited from the mother.

Regarding the brother's (B) genetic value contributed by the shared part, it is directly correlated with the genetic and environmental values of his sib-in-law (D). Therefore, the covariance between the brother's genetic value contributed by the shared part ($H_{1/2(shared)} + H_{3/4(shared)}$) and the sib-in-law's total genetic value (G_D) is:

$$Cov\left(H_{1/2(shared)} + H_{3/4(shared)}, G_D\right)$$

$$= Cov\left(H_{1/2(shared)}, G_D\right) + Cov\left(H_{3/4(shared)}, G_D\right)$$

$$= \frac{1}{2}Cov\left(H_{1/2}, G_D\right) + \frac{1}{2}Cov\left(H_{3/4}, G_D\right)$$

$$= \frac{1}{4} Cov(G_s, G_D) + \frac{1}{4} Cov(G_s, G_D)$$

$$= \frac{1}{2} Cov(G_s, G_D)$$

$$= \frac{1}{2} \rho h_{AM}^2 Var G_{AM} \qquad Eq15$$

And the covariance between the brother's genetic value contributed by the shared part ($H_{1/2(shared)} + H_{3/4(shared)}$) and the sib-in-law's environmental value (E_D) is:

$$Cov\left(H_{1/2(shared)} + H_{3/4(shared)}, E_D\right)$$

$$= Cov\left(H_{1/2(shared)}, E_D\right) + Cov\left(H_{3/4(shared)}, E_D\right)$$

$$= \frac{1}{2} Cov\left(H_{1/2}, E_D\right) + \frac{1}{2} Cov\left(H_{3/4}, E_D\right)$$

$$= \frac{1}{4} Cov(G_s, E_D) + \frac{1}{4} Cov(G_s, E_D)$$

$$= \frac{1}{2} Cov(G_s, E_D)$$

$$= \frac{1}{2} \rho \sqrt{h_{AM}^2 (1 - h_{AM}^2) Var G_{AM} Var E_{AM}} \qquad Eq16$$

Regarding the brother's (B) genetic value contributed by the non-shared part of paternal origin ($H^*_{1/2(unique)}$), it is correlated with his full-sibling's (S) genetic value of maternal origin ($H_{3/4}$) due to assortative mating between the father (F) and the mother (M). Similarly, $H^*_{3/4(unique)}$ is correlated with $H_{1/2}$.

Therefore, the brother's (B) genetic value contributed by the non-shared part is indirectly correlated with his sib-in-law's (D) genetic and environmental values through the genetic value of his full-sibling (S) and the covariance between the brother's genetic value contributed by the non-shared part ($H^*_{1/2(unique)} + H^*_{3/4(unique)}$) and the sib-in-law's total genetic value (G_D) is:

$$Cov\left(H^*_{1/2(unique)} + H^*_{3/4(unique)}, G_D\right)$$

$$= Cov\left(H^*_{1/2(unique)}, G_D\right) + Cov\left(H^*_{3/4(unique)}, G_D\right)$$

$$= \frac{1}{2}Cov\left(H^*_{1/2}, G_D\right) + \frac{1}{2}Cov\left(H^*_{3/4}, G_D\right)$$

$$= \frac{1}{2}Cov\left(\rho h^2_{AM} H_{3/4}, G_D\right) + \frac{1}{2}Cov\left(\rho h^2_{AM} H_{1/2}, G_D\right)$$

$$= \frac{1}{2}\rho h^2_{AM} Cov\left(H_{3/4}, G_D\right) + \frac{1}{2}\rho h^2_{AM} Cov\left(H_{1/2}, G_D\right)$$

$$= \frac{1}{4}\rho h^2_{AM} Cov(G_S, G_D) + \frac{1}{4}\rho h^2_{AM} Cov(G_S, G_D)$$

$$= \frac{1}{2}\rho h^2_{AM} Cov(G_S, G_D)$$

$$= \frac{1}{2}(\rho h^2_{AM})^2 Var G_{AM} \qquad Eq17$$

Similarly, the covariance between the brother's genetic value contributed by the non-shared part ($H^*_{1/2(unique)} + H^*_{3/4(unique)}$) and the sib-in-law's environmental value (E_D) is:

$$Cov\left(H^*_{1/2(unique)} + H^*_{3/4(unique)}, E_D\right)$$

$$= Cov\left(H^*_{1/2(unique)}, E_D\right) + Cov\left(H^*_{3/4(unique)}, E_D\right)$$

$$= \frac{1}{2}Cov\left(H^*_{1/2}, E_D\right) + \frac{1}{2}Cov\left(H^*_{3/4}, E_D\right)$$

$$= \frac{1}{2}Cov\left(\rho h^2_{AM} H_{3/4}, E_D\right) + \frac{1}{2}Cov\left(\rho h^2_{AM} H_{1/2}, E_D\right)$$

$$= \frac{1}{2}\rho h^2_{AM} Cov\left(H_{3/4}, E_D\right) + \frac{1}{2}\rho h^2_{AM} Cov\left(H_{1/2}, E_D\right)$$

$$= \frac{1}{4}\rho h^2_{AM} Cov(G_S, E_D) + \frac{1}{4}\rho h^2_{AM} Cov(G_S, E_D)$$

$$= \frac{1}{2}\rho h^2_{AM} Cov(G_S, E_D)$$

$$= \frac{1}{2}\rho^2 h^2_{AM}\sqrt{h^2_{AM}(1 - h^2_{AM}) Var G_{AM} Var E_{AM}} \qquad Eq18$$

Hence, by adding $Eq15$ and $Eq17$ and $Eq16$ and $Eq18$, it is possible to get the covariance between the brother's total genetic value (G_B) and the sib-in-law's total genetic value (G_D) and the covariance between the brother's total genetic value (G_B) and the sib-in-law's environmental value (E_D) respectively.

$$Cov\left(H_{1/2(shared)} + H_{3/4(shared)}, G_D\right) + Cov\left(H^*_{1/2(unique)} + H^*_{3/4(unique)}, G_D\right)$$

$$= Cov\left(H_{1/2(shared)} + H_{3/4(shared)} + H^*_{1/2(unique)} + H^*_{3/4(unique)}, G_D\right)$$

$$= Cov(G_B, G_D)$$

$$= \frac{1}{2}\rho h^2_{AM} VarG_{AM} + \frac{1}{2}(\rho h^2_{AM})^2 VarG_{AM}$$

$$= \frac{1 + \rho h^2_{AM}}{2}\rho h^2_{AM} VarG_{AM} \qquad Eq19$$

And

$$Cov\left(H_{1/2(shared)} + H_{3/4(shared)}, E_D\right) + Cov\left(H^*_{1/2(unique)} + H^*_{3/4(unique)}, E_D\right)$$

$$= Cov\left(H_{1/2(shared)} + H_{3/4(shared)} + H^*_{1/2(unique)} + H^*_{3/4(unique)}, E_D\right)$$

$$= Cov(G_B, E_D)$$

$$= \frac{1}{2}\rho\sqrt{h^2_{AM}(1 - h^2_{AM})VarG_{AM}VarE_{AM}} + \frac{1}{2}\rho^2 h^2_{AM}\sqrt{h^2_{AM}(1 - h^2_{AM})VarG_{AM}VarE_{AM}}$$

$$= \frac{1 + \rho h^2_{AM}}{2}\rho\sqrt{h^2_{AM}(1 - h^2_{AM})VarG_{AM}VarE_{AM}} \qquad Eq20$$

$Eq19$ and $Eq20$ demonstrate the genetic and genetics-by-environment correlation between FSIL, respectively. However, unlike the POIL situation, the brother's environmental value (E_B) is not correlated with his sibling's genetic value (G_S), and thus there is no environmental or environment-by-genetics correlation between FSIL. Hence,

$$Cov(E_B, E_D) = 0 \qquad Eq21$$

$$Cov(E_B, G_D) = 0 \qquad Eq22$$

By summing $Eq19$, $Eq20$, $Eq21$ and $Eq22$ up, it is possible to obtain the phenotypic covariance and correlation between FSIL

$$Cov(G_B, G_D) + Cov(G_B, E_D) + Cov(E_B, E_D) + Cov(E_B, G_D)$$

$$= Cov(G_B, G_D + E_D) + Cov(E_B, G_D + E_D)$$

$$= Cov(G_B + E_B, G_D + E_D)$$

$$= Cov(P_B, P_D)$$

$$= \frac{1 + \rho h_{AM}^2}{2} \rho h_{AM}^2 VarG_{AM} + \frac{1 + \rho h_{AM}^2}{2} \rho \sqrt{h_{AM}^2 (1 - h_{AM}^2) VarG_{AM} VarE_{AM}} + 0 + 0$$

$$= \frac{1 + \rho h_{AM}^2}{2} \rho h_{AM}^2 VarP_{AM} \qquad Eq23$$

And

$$Cor(P_B, P_D) = \frac{1 + \rho h_{AM}^2}{2} \rho h_{AM}^2 \qquad Eq24$$

Thus, under assortative mating, the expected phenotypic correlation between FSIL is $\frac{1+\rho h_{AM}^2}{2} \rho h_{AM}^2$, which is the product of the expected phenotypic correlation between full-siblings ($\frac{1+\rho h_{AM}^2}{2} h_{AM}^2$) and the expected phenotypic correlation between partners (ρ).

With ρ of 0.3 and h_{AM}^2 of 80%, the phenotypic correlation between sibs and sib-in-laws is 0.1488, much higher than the expected correlation of 0 under random mating.

I summarised the equations of expected resemblance between nuclear family members either obtained from the literature [188,196] or derived by myself in Table 5.1.

Table 5.1. Resemblance between different types of 1st degree relatives

Relationship	Correlation Type	Random Mating	Assortative Mating (Equilibrium)
Spouse and Spouse	$CorP$	0	ρ
	$CorG$	0	ρh_{AM}^2
	$CorE$	0	$\rho(1-h_{AM}^2)$
Parent and Offspring	$CorP$	$\frac{h_o^2}{2}$	$\frac{1+\rho}{2}h_{AM}^2$
	$CorG$	0.5	$\frac{1+\rho h_{AM}^2}{2}$
	$CorE$	0	0
Parent and Offspring-in-Law	$CorP$	0	$\frac{1+\rho}{2}\rho h_{AM}^2$
	$CorG$	0	$\frac{1+\rho h_{AM}^2}{2}\rho h_{AM}^2$
	$CorE$	0	$\frac{\rho^2 h_{AM}^2(1-h_{AM}^2)}{2}$
Full Sibling and Full Sibling	$CorP$	$\frac{h_o^2}{2}$	$\frac{1+\rho h_{AM}^2}{2}h_{AM}^2$
	$CorG$	0.5	$\frac{1+\rho h_{AM}^2}{2}$
	$CorE$	0	0
Full Sibling and Full-Sibling-in-Law	$CorP$	0	$\frac{1+\rho h_{AM}^2}{2}\rho h_{AM}^2$
	$CorG$	0	$\frac{1+\rho h_{AM}^2}{2}\rho h_{AM}^2$
	$CorE$	0	0

5.3 Simulation Study

To examine the performance of these formulae listed in Table 5.1 in a range of different scenarios, I conducted a simulation study.

5.3.1 Simulated Population Structure

I simulated 8 assortative mating cohorts, 21 generations (the number included the founder populations) for each cohort and 100k individuals (50k males and 50k females) for each generation. All individuals were assortatively mated and produced one offspring of each sex. There was no cross-generation mating, full-sibling mating or first-cousin mating.

For details about how assortatively mated individuals were simulated see Section 5.3.3.

5.3.2 Simulated Trait Architecture

Two thousand random variables were generated from a binomial distribution $\mathcal{B}(1, 0.5)$ (0 or 1 with equal probability) to be the haplotypes of one thousand causal loci in the founder populations. Thus, all causal loci were unlinked in the founder populations with allele frequencies around 0.5. Each locus acted independently when being transmitted from parents to offspring, i.e. a recombination rate of 50%, and consequently, in the absence of assortative mating, loci should remain in LE in any subsequent generations.

Each locus was assigned an effect size for the reference allele (the allele coded as 1). Effect sizes were generated from a normal distribution $\mathcal{N}(0, \frac{2VarG_o}{N_{loci}})$ and were constant over generations. The genetic value of an individual was calculated as $G = ga$, where G is the genetic value, g is the array of genotype, a is the vector of effect sizes. Thus, the expectation of additive genetic variance equals $\sum 2p_i q_i a_i^2 = VarG_o$ for the founder populations of assortative mating cohorts.

Each individual was assigned an environmental effect (E), derived from a normal distribution $\mathcal{N}(0, VarP_o - VarG_o)$. The distribution of environmental effects was constant over generations and therefore $VarE = VarE_o = VarP_o - VarG_o$.

Finally, the phenotype (P) for any individual was calculated as $P = G + E$. The phenotype should follow multivariate normal distribution $MV\mathcal{N}(0, VarP_o)$ and $VarP_o$ was always set to 1. For the founder populations of assortative mating cohorts, $VarP = VarP_o$, whereas for progeny generations of populations under assortative mating, $VarP$ should be larger than $VarP_o$ because of the increase in additive genetic variance caused by assortative mating.

Table 5.2 shows the parameters used for data simulation.

Table 5.2. Parameters for simulated cohorts

Cohort	Mating Type	ρ_0	$VarG_0$	$VarE_0$	$VarP_0$	N_{loci}
1	Assortative Mating	0.15	0.20	0.80	1	1000
2			0.40	0.60		
3			0.60	0.40		
4			0.65	0.35		
5		0.3	0.20	0.80		
6			0.40	0.60		
7			0.60	0.40		
8			0.65	0.35		

5.3.3 Procedure of Assortative Mating

To generate assortment in mate choice, another 50k intermediate values were simulated for each generation, in addition to the phenotypes simulated in Section 5.3.2, The steps used to obtain assortatively mated simulated individuals are given below:

1. Pick a simulated male and generate an intermediate value for that male by selecting a random value from the normal distribution (mean = male's simulated phenotype, s.d. = sigma). Sigma is a parameter associated with couple correlation and phenotypic variance (see below)
2. Repeat Step 1 for all 50k simulated males and get 50k intermediate values.
3. Rank 50k females' phenotypes simulated in Section 5.3.2 and rank 50k intermediate values generated in Step 1 and Step 2.
4. Match the rank of males' intermediate values and the rank of females' simulated phenotypes and mate the corresponding individuals with the same rank.
5. For example:
    a. ID 1 is a male with simulated phenotype of 2
    b. Randomly select an intermediate value for ID 1 from normal distribution (mean=2, s.d.=sigma) and the value is 1.8
    c. 1.8 is the 102nd largest values among all intermediate values
    d. The 102nd largest simulated phenotype in female is 1.83 and belongs to female ID 7312
    e. Mate male ID1 and female ID 7312
    f. Repeat a to e for the remaining male samples

I found that the correlation between males' simulated phenotypes and their intermediate values is identical to the correlation between males' simulated phenotypes and their chosen spouses' simulated phenotypes mated by this way, e.g. same ranking provides similar correlation.

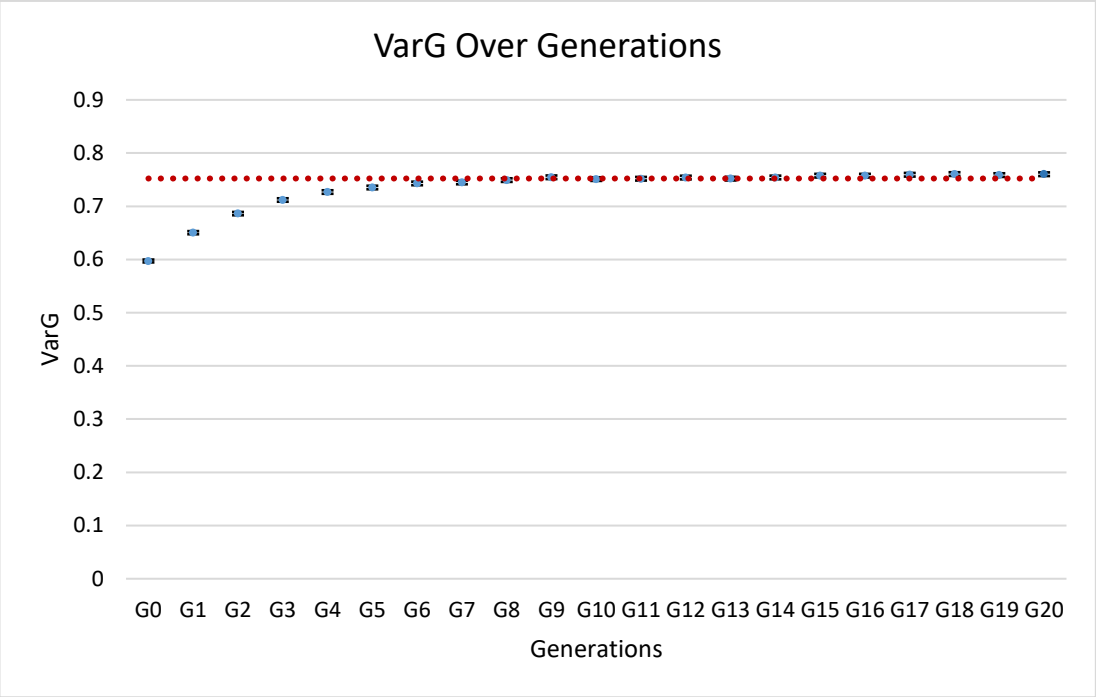
The next step is to choose the right parameter, sigma, to make intensity of assortative mating (spousal phenotypic correlation) ρ equivalent to 0.15 or 0.3. With a fixed $VarP$ of 1, the empirical distribution for the sigma and spousal phenotypic correlation is given in Figure S5.1. Based on Figure S5.1, when sigma = 3.2 and 6.5, ρ = 0.3 and

0.15 respectively. However, $VarP$ increases over generations till populations reach equilibrium. Thus, the observed spousal phenotypic correlation ρ might differ slightly from the expectation.

5.3.4 Equilibrium of Assortative Mating

In my simulation study, there were 1k causal loci for a trait and 100k individuals per generation. This large sample size made populations approach equilibrium very quickly. I plot the changes in genetic variance over generations for one simulated assortative mating cohort in Figure 5.3.

Figure 5.3 An example of how genetic variance (VarG) changes over generations for populations under assortative mating.



The example is from a simulated assortative mating cohort with $VarG_0 = 0.6$ and $\rho_0 = 0.3$ ($\rho \approx 0.31$ when the populations reached equilibrium). The red line is the expectation of genetic variance in populations at equilibrium of assortative mating calculated as $\frac{VarG_0}{1-\rho h^2_{AM}}$, which is ~0.75.

Figure 5.3 shows that, based on parameters used in my simulation, 8-10 generations of assortative mating are enough for the simulated population to reach equilibrium of assortative mating as the plot is flat and the estimates of genetic variance are no longer significantly different from each other after 8-10 generations. Therefore, the last 10 generations of this simulated cohort have reached equilibrium of assortative mating.

5.3.5 Resemblance between Nuclear Family Members

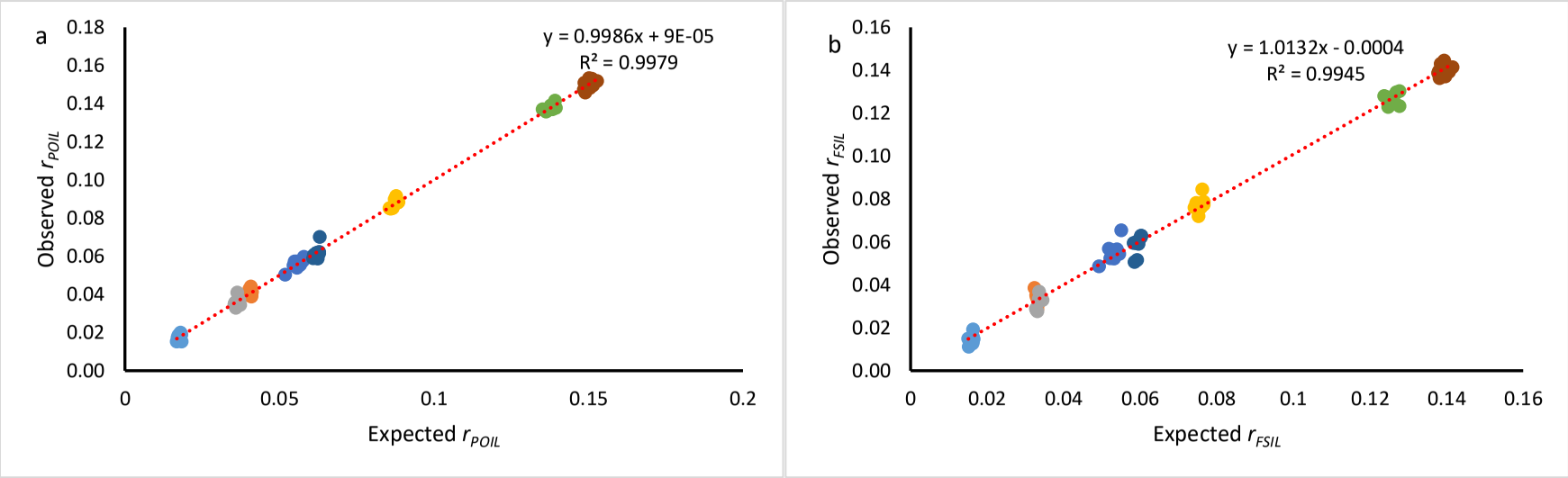
To test whether the equations of expected resemblances between relatives listed in Table 5.1 are accurate, I compared the observed values estimated from simulated data against the expected values calculated based on formulae in Table 5.1 by regression.

Two examples of the phenotypic correlation between POIL and between FSIL are given in Figure 5.4.

Regressing the observed values on the expected values, both the regression coefficients and the variance explained by the regression (R^2) are close to 1, which shows agreement between the formulae derived and the results from the simulation study for these two relationships.

Analogously, all equations listed in Table 5.1 were verified by linear regression (Figure S5.2). However, if the expected correlation between relatives is 0 rather than positive (e.g. the expected environmental correlation between parents and offspring is 0), I conducted sign test to test whether the observed values significantly deviate from 0 instead.

Figure 5.4. Observed vs expected phenotypic correlation between parents and offspring-in-law (r_{POIL}) and between full-siblings and siblings-in-law (r_{FSIL})



Red line: regression line; R^2: variance explained by the regression; The expected values of r_{POIL} and r_{FSIL} are calculated as $\frac{1+\rho}{2}\rho h^2_{AM}$ and $\frac{1+\rho h^2_{AM}}{2}\rho h^2_{AM}$, respectively; Each coloured full circle in the legend below refers to one simulated population at equilibrium with the following founder genetic variances ($VarG_0$) and assortative mating intensities (ρ_0):

$VarG_0$	$\rho_0 = 0.15$	$\rho_0 = 0.30$
0.20	●	●
0.40	●	●
0.60	●	●
0.65	●	●

5.4 Novel Pedigree Studies using in-Law Relatives to Estimate Heritability and Intensity of Assortative Mating

By using simulated data, I confirmed that the expected resemblances between nuclear family members under assortative mating listed in Table 5.1 are accurate. Therefore, by using two types of listed relationships, I could estimate the heritability (h^2_{AM}) and the intensity of assortative mating (ρ) for a trait as the expected phenotypic correlations between nuclear family members are linked to h^2_{AM} and ρ.

However, there are confounding factors that could affect the resemblances between relatives. For example, the couple correlation (r_{CP}) might be influenced by common couple environment; the parent-offspring correlation (r_{PO}) might be influenced by common family environment; and the full-sibling correlation (r_{FS}) might be inflated by dominance and epistasis as well as common rearing environment.

But, the correlation between parent- and offspring-in-law (POIL), r_{POIL}, and between sib- and sib-in-law (FSIL), r_{FSIL}, is not affected by dominance, epistasis, household effect and common environment shared between partners or siblings. Therefore, by using the phenotypic correlations of two different types of in-law relationship or a type of in-law relationship and a type of biological relationship, I could estimate ρ and h^2_{AM} with less bias from the confounding factors mentioned above for populations at equilibrium of assortative mating.

For example, based on the expected resemblance listed in Table 5.1, I could estimate ρ and h^2_{AM} by using POIL and FSIL relationships.

$$\rho = \frac{r_{POIL} - 2r_{FSIL} + \sqrt{r^2_{POIL}(1+8r_{FSIL})}}{2r_{FSIL}} \qquad Eq25$$

$$h^2_{AM} = \frac{2r_{POIL}}{\rho(1+\rho)} \qquad Eq26$$

Regarding parent-offspring and POIL relationships, the equations are as follows:

$$\rho = \frac{r_{POIL}}{r_{PO}} \qquad Eq27$$

$$h^2_{AM} = \frac{2r_{PO}}{1+\rho} \qquad Eq28$$

Similarly, for sib-sib and FSIL relationships,

$$\rho = \frac{r_{FSIL}}{r_{FS}} \qquad Eq29$$

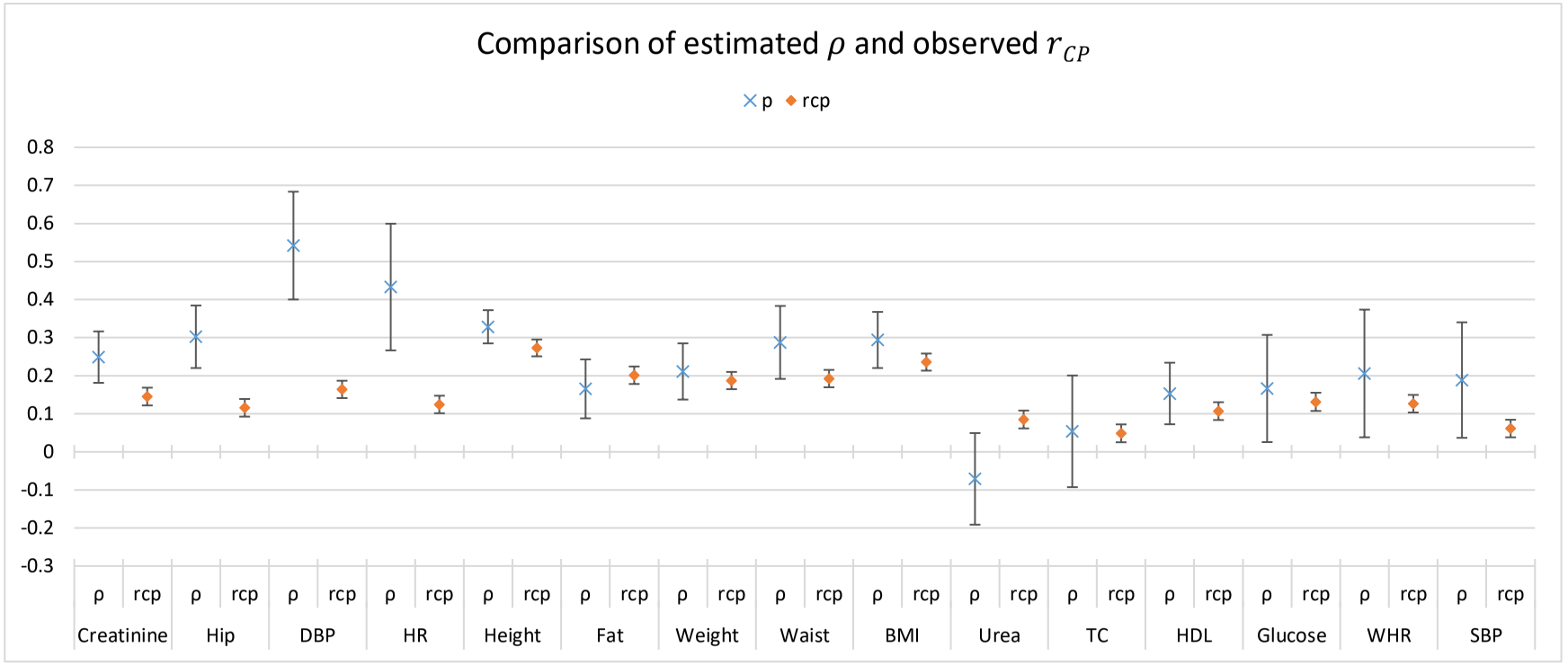
$$h^2_{AM} = \frac{-1+\sqrt{1+8r_{FSIL}}}{2\rho} \qquad Eq30$$

I examined equations $Eq25$ to $Eq30$ using simulated data (Figure S5.3). The overall estimates of ρ and h^2_{AM} are unbiased, but the standard errors of the estimate from POIL-FSIL design ($Eq25$ and $Eq26$) are large, especially for traits with low founder heritability and low intensity of assortative mating, e.g. the 1st simulated cohort ($VarG_0 = 0.2$ and $\rho = 0.15$) in Table 5.2.

Subsequently, I conducted a novel pedigree study using related individuals in GS20K. First, I checked how many in-law relationships are there. Unfortunately, although GS20K contains deep pedigrees, it does not have a lot of pairs of different types of in-law relationship. The highest number of in-law relationships is FSIL. In GS20K, there is ~1.9k pairs of full-siblings for which one's spouse happened to also participate in the study; whereas the number of informative POIL pairs is low, less than 300. Consequently, I conducted a sib-sib and FSIL pedigree study to estimate the heritability and the intensity of assortative mating (if any) for anthropometric and cardio-metabolic traits in GS20K using $Eq29$ and $Eq30$. Phenotypes have been adjusted for sex, age and age^2 and sex-by-age interaction.

Based on $Eq29$, I predicted the intensity of assortative mating ρ using observed phenotypic correlations between full-siblings (r_{FS}) and between FSIL (r_{FSIL}). Afterwards, I compared the predicted ρ to the observed phenotypic correlation between partners (r_{CP}) for the same trait in Figure 5.5 (for detailed values see Table S5.1) to see whether they differ. A significant difference between ρ and r_{CP} would imply that there is shared couple environment, apart from assortative mating.

Figure 5.5 Predicting the intensity of assortative mating (ρ) using sib-sib and sib-in-law relationships and comparing the predicted intensity to the observed phenotypic correlation between partners (r_{CP}) in GS:SFHS



Standard errors of ρ were estimated using delta method, for detail sees Text S5.1

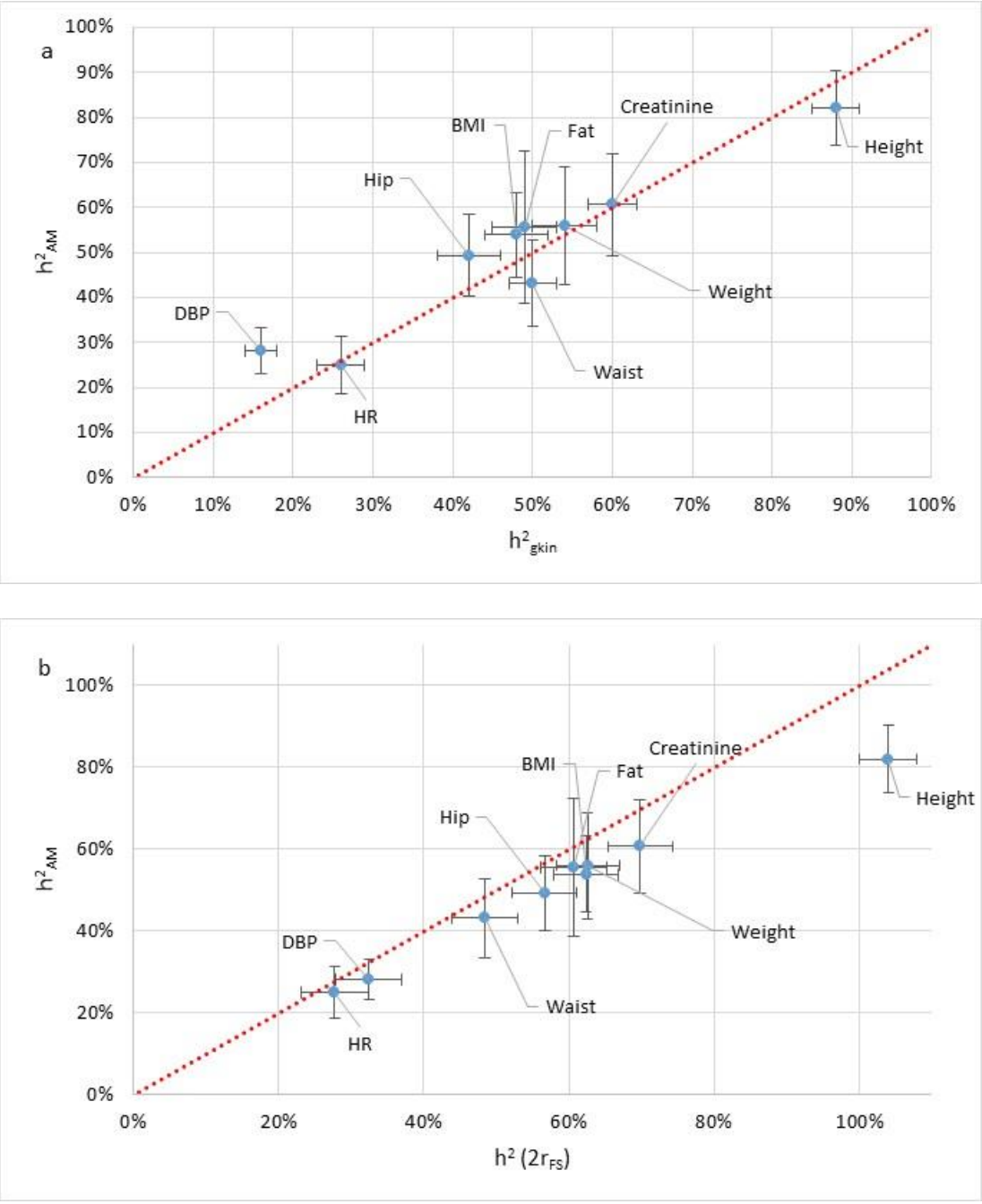
Results indicate that (Figure 5.5), for creatinine, hip circumference, diastolic blood pressure (DBP) and heart rate (HR), there is evidence of assortative mating as the estimates of ρ are significantly larger than 0; and there is evidence that the couple similarity is not only contributed by assortative mating as the estimated ρ is significantly larger than the observed r_{CP} for the same trait. Regarding height, fat, weight, waist circumference and BMI, there is significant evidence of assortative mating but the estimated ρ is not significantly different from observed r_{CP} for the same trait, which suggests that the couple similarity is caused by assortative mating solely; or the difference is not large enough to be detected using my real data. For urea, TC, HDL, glucose, waist-to-hip ratio (WHR) and systolic blood pressure (SBP), there is no evidence of assortative mating. However, the observed r_{CP} is significant for these traits, indicating that the couple similarity is caused by shared couple environment solely.

For traits showing evidence of assortative mating, I estimated the heritability h^2_{AM} using the observed phenotypic correlation between full-siblings (r_{FS}) and the predicted intensity of assortative mating (ρ) based on $Eq30$ (Table S5.1). Afterwards, I compared the estimates of h^2_{AM} to the estimates of h^2_{gkin} (h^2_g+h^2_{kin}, from the selected model in Table 2.3) obtained in Chapter 2 for the same trait to see how well the estimates of heritability given by two different methods agree with each other.

Figure 5.6a shows that the estimates of h^2_{AM} obtained from sib-sib and FSIL relationships correspond quite well with h^2_{gkin} estimates from Chapter 2 as estimates are similar and no significant difference between h^2_{AM} and h^2_{gkin} is observed for the same trait, except for DBP for which h^2_{AM} is higher than h^2_{gkin}.

Subsequently, I compared the estimates of h^2_{AM} to twice the observed phenotypic correlation between full-siblings because $2r_{FS}$ is the heritability estimate from sib-sib relationship assuming random mating. As shown in Figure 5.6b, all h^2_{AM} estimates are lower than $2r_{FS}$, which indicates that estimating heritability from sibling relationship without considering the assortment in mate choice leads to inflated heritability estimates for traits under assortative mating. The inflation for height is significant, although this is not the case for other traits in my data.

Figure 5.6 Estimating h^2_{AM} using predicted ρ and observed r_{FS} based on $Eq30$ and comparing estimates of h^2_{AM} to estimates of h^2_{gkin} and $2r_{FS}$.



Horizontal and vertical bars show standard errors of h^2_{gkin} (or $2r_{FS}$ in plot b) estimates and h^2_{AM} estimates respectively. The standard errors of h^2_{AM} estimates were estimated based on delta method, for detail sees Text S5.1.

5.5 Conclusion and Discussion

In this study, I mathematically derived the expected resemblance between sib-sib-in-law (FSIL, the relationship between one's full-siblings and one's spouse) and between parent-offspring-in-law (POIL, the relationship between one's parents and one's spouse) under assortative mating. I found that, under assortative mating, the expected resemblance between FSIL and POIL are positive and relate to the heritability (h^2_{AM}, the heritability for populations reached equilibrium at assortative mating) and the intensity of assortative mating (ρ, how strong the assortment in mate choice is). This finding was confirmed by simulation study (Figure 5.4).

I developed novel methods to estimate heritability and the intensity of assortative mating by using the resemblances between in-law relatives in complex pedigrees, such as sib-sib and FSIL relationships ($Eq29$ and $Eq30$), parent-offspring and POIL relationships ($Eq27$ and $Eq28$) and POIL and FSIL relationships ($Eq25$ and $Eq26$). These equations were also validated by simulation study (Figure S5.3).

Subsequently, I conducted sib-sib and FSIL studies to estimate h^2_{AM} and ρ for anthropometric and cardio-metabolic traits in GS20K as FSIL is the most frequent in-law relationship in GS20K (~1.9k pairs). Results indicate that, there is some evidence of assortative mating for 9 out of 15 traits investigated in this study (Figure 5.5) and for traits having evidence of assortative mating the estimates of h^2_{AM} are similar with the estimates of h^2_{gkin} from Chapter 2 (Figure 5.6a), which suggests that this novel pedigree study performs reasonably well.

However, my estimates of ρ and h^2_{AM} from sib-sib and FSIL relationships might be influenced by confounding factors shared among full-siblings and FSIL. For example, observed similarity between full-siblings (r_{FS}) might be larger than expectation due to dominance, epistasis and shared environment and thus resulting in underestimation of ρ and overestimation of h^2_{AM}. This might explain why the estimate of h^2_{AM} is significantly larger than the estimate of h^2_{gkin} for DBP for which sibling environment seems to contribute a large proportion of the sibling similarity (sibling environment explains 8% of total phenotypic variance for DBP, Table 2.3). Another potential confounder is regional effects, e.g. the observed similarity between FSIL (r_{FSIL}) might

be larger than expectation due to living in the same region or FSIL are genetically more similar compared to two random individuals due to coming from the same region. Regional effects will lead to overestimation of ρ and underestimation of h^2_{AM}. Furthermore, overestimation of ρ might result in false positive estimates of assortative mating.

Regarding creatinine, hip circumference, DBP and HR, there is significant evidence that the observed similarity between partners is different from the predicted ρ which suggests that shared common couple environment after cohabiting might also contribute to the couple similarity. This gives us an opportunity to dissect the observed couple similarity into the effects of shared common couple environment due cohabitation and the effects of assortative mating. However, in this study, the estimates of ρ are significant larger than the observed r_{CP}. This indicates that, if there is shared couple environment, then the effects of shared couple environment should be in the opposite direction to the effects of assortative mating, i.e. making couples dissimilar. It is possible that there are other potential undetected cofounding factors that inflates r_{FSIL} but not r_{FS} (or inflates r_{FSIL} greater than r_{FS}) which leads to overestimation of ρ for these traits. Further investigation is required.

In the future, I plan to conduct pedigree study to estimate heritability and the intensity of assortative mating using POIL and FSIL relationships in UK biobank as there might be a sufficient number of POIL and FSIL pairs to provide reasonable estimates. The advantage of using two types of in-law relationships is that, there is no dominance, epistasis, household effect, common couple environment and common sibling environment shared by in-law relatives other than partners. Therefore, for traits under assortative mating, a pedigree study using two types of in-law relationships might provide more accurate estimates of ρ and h^2_{AM}, mitigating the bias in estimation of heritability in traditional pedigree study such as MZ-DZ twin study and parent-offspring regression caused by the confounding factors shared among nuclear family members. Future research might also explore the possibility of using mixed model and likelihood approaches to estimate heritability, assortative mating and family environment effects, as well as dominance and some other effects, simultaneously using resemblances between all available relatives in deep pedigrees.

Chapter 6: Conclusions and Future Work

In this thesis, I conducted variance component analyses (Chapter 2), GWAS (Chapter 3), a prediction study (Chapter 4) and an assortative mating study (Chapter 5) to explore the trait architecture for anthropometric and cardio-metabolic traits, including height, weight, fat, body mass index (BMI), hip circumference, waist circumference, waist-to-hip ratio (WHR), a body shape index (ABSI), levels of creatinine, urea, total cholesterol (TC) and high-density lipoprotein (HDL) in serum, levels of glucose in blood after a four hour fasting period, systolic and diastolic blood pressure and heart rate. The data analysed were collected from a cohort made up of ~20k individuals of recent Scottish descent genotyped for over 520k common SNPs across the genome.

In Chapter 2, I conducted variance component analyses to identify the sources of trait variation using genomic relationship matrices (GRM) and similarity matrices under a REML framework. I discovered that, for most traits investigated, the major contributors to trait variation were SNP-associated genetic effects, pedigree-associated genetic effects, couple effects and sibling effects (Table 2.3).

SNP-associated genetic effects refer to genetic effects captured by common variants inherited from distant ancestors that are associated with genotyped SNPs at the population level, whereas pedigree-associated genetic effects represent additional genetic effects due to genetic variants that segregate within pedigrees but are not associated with genotyped SNPs at the population level, such as rare variants, CNVs and other structural variants, that are captured due to strong linkage in high-order pedigrees. On average, SNP- and pedigree-associated genetic effects each explain ~50% of the genetic variance (Figure 2.4) and the total heritability (sum of two genetic effects) matches that published in twin studies (Table 2.4), which indicates little heritability is missing. Future work could focus on identifying the genetic variants contributing to pedigree-associated genetic effects by conducting GWAS in isolated populations in which the frequencies of those genetic variants could be higher.

In GS:SFHS, the average age of participants is around 50 years which suggests that individuals currently sharing a common household environment will largely be

couples. Therefore, in this study, sibling effects and couple effects refer to the environmental effects due to past rearing environment shared by full-siblings and due to current environment shared by partners, respectively. However, couple effects also involve assortative mating. Assortative mating increases the similarity between partners and thus is confounded with the effects contributed by the shared environment.

In order to identify the presence of assortative mating, in Chapter 5, I conducted an assortative mating study. I developed a novel method that uses the relationships between in-law relatives to determine whether the observed phenotypic correlation between partners is contributed by environment or assortative mating or a mixture of both as well as to estimate the heritability for the range of traits studied. Using this novel method, I found significant evidence of assortative mating for creatinine, hip circumference, diastolic blood pressure, heart rate, height, fat, weight, waist circumference and BMI and for the first 4 traits, there is significant evidence that the observed phenotypic correlation between partners is not only contributed by assortative mating (Figure 5.5). The heritability estimates given by this approach are close to those obtained in Chapter 2 (Figure 5.6). In the future, the next step is to separate the observed phenotypic correlation between partners into the effects due to assortment in mate choice and the effects due to shared environment as a result of the cohabitation. To solve that, an assumption about the relationship between assortative mating and shared couple environment is required, that is for an individual whether the effect due to assortative mating is correlated with the effect due to shared environment or they are independent. Although assortative mating is a mate choice at the phenotype level which is blind to genetics and environmental factors, it generates both genetic and environmental correlations between partners. Since one's lifestyle and habits would not change completely across time, there might be a positive correlation between the environmental effects unknowingly shared by partners due to mate choice prior to cohabitation and the environmental effects shared by partners after cohabitation due to common living environment. This enhances the difficulty of dissecting the observed phenotypic correlation between partners into different sources.

In Chapter 3, I conducted GWAS to detect genetic variants attributable to SNP-associated genetic effects. However, unlike traditional GWAS study, my extended GWAS model considers pedigree-associated genetic effects, couple effects and sibling

effects in addition to SNP-associated genetic effects, i.e. the factors shown in Chapter 2 that contribute to trait variation. This approach could remove the false positive associations due to genetics-by-environment correlation shared between relatives and increase detection power by providing smaller standard errors for the estimates of SNP effect sizes. By a comprehensive GWAS performance comparison, I provided evidence that, in general, the extended method provides lower FDR and higher detection power for traits investigated compared to the traditional method that only accounts for SNP-associated genetic effects. There are, however, some exceptions. There is significant evidence that the traditional method works better for height and creatinine (Figure 3.1 and Table 3.3). Note, there is also some evidence that the traditional method works better for heart rate and systolic blood pressure (Figure 3.1 and Table 3.2), but that is because most signals detected are false positives (Table S3.2). A plausible hypothesis is that, as mentioned above, assortative mating generates positive genetic and environmental correlations between members of a couple, i.e. ρh^2 for genetic correlation and $\rho(1-h^2)$ for environmental correlation. Compared to the traditional method which does not model couple effects, for traits under assortative mating, modelling couple effects in the extended GWAS method could remove the extra genetic variance shared by partners due to assortative mating (ρh^2 part), leading to lower detection power; simultaneously, modelling couple effects in the extended GWAS method could remove the environmental correlation between partners due to assortative mating ($\rho(1-h^2)$ part), leading to higher detection power. Therefore, leaving alone the influence of shared common environment, for traits under assortative mating, whether this extended GWAS method benefits from modelling couple effects or not perhaps depends on the difference in gain and loss of detection power, i.e. the magnitude of ρh^2 and $\rho(1-h^2)$. For height and creatinine, there is evidence of assortative mating detected in Chapter 5 (Figure 5.5) and ρh^2 is much larger than $\rho(1-h^2)$ (Table S5.1), which leads to lower detection power for the extended method compared to the traditional method. On the contrary, although there is evidence of assortative mating for waist circumference, fat, weight and BMI (Figure 5.5), the extended method outperformed the traditional method (Figure 3.1, Table 3.2 and Table 3.3) as ρh^2 is no more than $\rho(1-h^2)$ (Table S5.1) for these traits. This hypothesis requires further investigation theoretically or by a simulation study.

In Chapter 4, I conducted a prediction study to predict the phenotypic values for some obesity-related traits including height, BMI, hip circumference, HDL and TC. Similar to the GWAS, my prediction model also includes pedigree-associated genetic effects, couple effects and sibling effects in addition to SNP-associated genetic effects. This makes my study different from traditional genomic prediction studies in which only SNP-associated genetic effects are included. Results indicate that, the prediction accuracy from the extended prediction method is, on average, ~1.6% higher for these traits (although these increases are non-significant), compared to the traditional prediction method (Table 4.1). Considering the fact that the prediction study was conducted using GS10K data in which there is a much lower proportion of distant, couple and sibling relationships compared to GS20K (Table 2.1) and that not everyone with appropriate relatives in the data could benefit from using the extended model unless his/her relatives are in the training set while the individual to predict remains in the validation set simultaneously, such non-significant increase in the prediction accuracy driven by limited number of corresponding individuals is still promising. By examining the prediction accuracy of particular subpopulations, I discovered that the prediction accuracy of individuals with relatives in the data is much higher than that of individuals without (Figure 4.3), e.g. for HDL, the prediction accuracy of individuals who have distant and sibling relationships in the data is over 40% whereas that of individuals who are unrelated to anyone else in the data is less than 15%. This points out that future prediction study could focus on maximising the prediction accuracy for each individual by using different prediction models based on what types of relative that individual has in the data.

To conclude, in this thesis, I have conducted studies to explore the trait architecture for anthropometric and cardio-metabolic traits. Owing to the deep and complex relationships in GS:SFHS, I am able to study the influence of familial genetic and environmental effects on trait variation, which is novel compared to most published trait architecture studies. Based on my observation, couple effects, which include both assortative mating and shared environment, are quite important for traits related to anthropometrics and cardio-metabolism. Disentangling how assortative mating and shared couple environment influence the architecture of these traits requires further investigation.

References

1. Mendel G. (1866) Versuche über Plflanzenhybriden [Experiments concerning plant hybrids, translated by William Bateson]. *Verhandlungen des naturforschenden Vereines in Brünn [Proceedings of the Natural History Society of Brünn]* 1865: 3-47.
2. Gerstein M.B., Bruce C., Rozowsky J.S., Zheng D., Du J., et al. (2007) What is a gene, post-ENCODE? History and updated definition. *Genome Research* 17 (6): 669-681.
3. Darwin C. (1859) *The Origin of Species by Means of Natural Selection; or the Preservation of Favoured Races in the Struggle for Life.* London: John Murray.
4. Darwin C. (1868) *The Variation of Animals and Plants under Domestication*. London: John Murray.
5. Roll-Hansen N. (1989) The crucial experiment of Wilhelm Johannsen. *Biology and Philosophy* 4 (3): 303-329.
6. de Vries H. (1889) *Intracellular Pangenesis*. Gager CS, translator; Gager CS, editor. Chicago: The Open Court Publishing Co. .
7. Morgan T.H., Sturtevant A.H., Muller H.J., Bridges C.B. (1915) *The Mechanism of Mendelian Heredity*. New York: Henry Holt.
8. Haldane J.B.S., Sprunt A.D., Haldane N.M. (1915) Reduplication in mice (Preliminary communication). *Journal of Genetics* 5 (2): 133–135.
9. Fisher R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Philosophical Transactions of the Royal Society of Edinburgh* 52: 399–433.
10. Visscher P.M., Hill W.G., Wray N.R. (2008) Heritability in the genomics era--concepts and misconceptions. *Nature Reviews Genetics* 9 (4): 255-266.
11. Wright S. (1921) Correlation and causation. *Journal of Agricultural Research* 20: 557-585.
12. Wright S. (1922) Coefficients of inbreeding and relationship. *The American Naturalist* 56 (645): 330-338.
13. Wright S. (1929) The evolution of dominance. *The American Naturalist* 63 (689): 556-561.

14. Polderman T.J., Benyamin B., De Leeuw C.A., Sullivan P.F., Van Bochoven A., et al. (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* 47 (7): 702-709.
15. Liu C., Dupuis J., Larson M.G., Cupples L.A., Ordovas J.M., et al. (2015) Revisiting heritability accounting for shared environmental effects and maternal inheritance. *Human Genetics* 134 (2): 169-179.
16. Rózsa J., Strand T.M., Montadert M., Kozma R., Höglund J. (2015) Effects of a range expansion on adaptive and neutral genetic diversity in dispersal limited Hazel grouse (Bonasa bonasia) in the French Alps. *Conservation Genetics* 17 (2): 401-412.
17. Avery O.T., Macleod C.M., Mccarty M. (1944) Studies on the chemical nature of the substance including transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *The Journal of Experimental Medicine* 79 (2): 137-158.
18. Watson J.D., Crick F.H.C. (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 171 (4356): 737-738.
19. Brenner S., Jacob F., Meselson M. (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190: 576 - 581.
20. Crick F.H., Barnett L., Brenner S., Watts-Tobin R.J. (1961) General nature of the genetic code for proteins. *Nature* 192: 1227-1232.
21. Nirenberg M., Leder P., Bernfield M., Brimacombe R., Trupin J., et al. (1965) RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proceedings of the National Academy of Sciences of the United States of America* 53 (5): 1161-1168.
22. Söll D., Ohtsuka E., Jones D.S., Lohrmann R., Hayatsu H., et al. (1965) Studies on polynucleotides, XLIX. Stimulation of the binding of aminoacyl-sRNA's to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids. *Proceedings of the National Academy of Sciences of the United States of America* 54 (5): 1378-1385.

23. Fiers W., Contreras R., de Wachter R., Haegeman G., Merregaert J., et al. (1971) Recent progress in the sequence determination of bacteriophage MS2 RNA. *Biochimie* 53 (4): 495-506.
24. Chow L.T., Gelinas R.E., Broker T.R., Roberts R.J. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12 (1): 1-8.
25. Berget S.M., Moore C., Sharp P.A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* 74 (8): 3171-3175.
26. Grodzicker T., Williams J., Sharp P., Sambrook J. (1974) Physical mapping of temperature-sensitive mutations of adenoviruses. *Cold Spring Harbor Symposia on Quantitative Biology* 39: 439-446.
27. Chang C., Bowman J.L., Dejohn A.W., Lander E.S., Meyerowitz E.M. (1988) Restriction fragment length polymorphism linkage map for Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* 85 (18): 6856-6860.
28. Paterson A.H., Lander E.S., Hewitt J.D., Peterson S., Lincoln S.E., et al. (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335 (6192): 721-726.
29. Botstein D., White R.L., Skolnick M., Davis R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32 (3): 314-331.
30. Nakamura Y., Leppert M., O'connell P., Wolff R., Houm T., et al. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235 (4796): 1616-1622.
31. Powell W., Machray G.C., Provan J. (1996) Polymorphism revealed by simple sequence repeats. *Trends in Plant Science* 1 (7): 215-222.
32. Röder M.S., Korzun V., Wendehake K., Plaschke J., Tixier M.H., et al. (1998) A microsatellite map of wheat. *Genetics* 149 (4): 2007-2023.
33. Temnykh S., Park D.W., Ayres N., Cartinhour S., Hauck N., et al. (2000) Mapping and genome organization of microsatellite sequences in rice (Oryza sativa L.). *Theoretical and Applied Genetics* 100 (5): 697-712.

34. International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431 (7011): 931-945.
35. NCBI (2016) Human genome overview. URL: http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/. Accssed on 09-August, 2016
36. The International HapMap Consortium (2003) The International HapMap Project. *Nature* 426: 789-796.
37. The International HapMap 3 Consortium (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467 (7311): 52-58.
38. Smith T., Vihinen M., Human Variome Project (2015) Standard development at the Human Variome Project. *Database : the journal of biological databases and curation* 2015: bav024.
39. Ring H.Z., Kwok P.Y., Cotton R.G.H. (2006) Human Variome Project: an international collaboration to catalogue human genetic variation. *Pharmacogenomics* 7 (7): 969-972.
40. The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature* 467 (7319): 1061-1073.
41. The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526 (7571): 68-74.
42. Genomics England (2016) The 100,000 Genomes Project. URL: https://www.genomicsengland.co.uk/the-100000-genomes-project/. Accssed on 09-August, 2016
43. Amberger J.S., Bocchini C.A., Schiettecatte F., Scott A.F., Hamosh A. (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Research* 43 (Database issue): D789-798.
44. OMIM (2016) OMIM gene map statistics. URL: http://www.omim.org/statistics/geneMap. Accssed on 15-August, 2016
45. The International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409 (6822): 928-933.

46. NCBI (2016) SNP summary. URL: http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi. Accssed on 09-August, 2016

47. Gray I.C., Campbell D.A., Spurr N.K. (2000) Single nucleotide polymorphisms as tools in human genetics. *Human Molecular Genetics* 9 (16): 2403-2408.

48. UK Biobank (2016) Genetic data. URL: http://www.ukbiobank.ac.uk/scientists-3/genetic-data/. Accssed on 09-August, 2016

49. Vanraden P.M. (2007) Genomic measures of relationship and inbreeding. *Interbull Bull* 37: 33-36.

50. Vanraden P.M. (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91 (11): 4414-4423.

51. Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 42 (7): 565-569.

52. Yang J., Lee S.H., Goddard M.E., Visscher P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* 88 (1): 76-82.

53. Haines J.L., Hauser M.A., Schmidt S., Scott W.K., Olson L.M., et al. (2005) Complement factor H variant increases the risk of age-related macular degeneration. *Science* 308 (5720): 419-421.

54. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447 (7145): 661-678.

55. Visscher P.M., Brown M.A., Mccarthy M.I., Yang J. (2012) Five years of GWAS discovery. *American Journal of Human Genetics* 90 (1): 7-24.

56. Zeggini E., Ioannidis J.P.A. (2009) Meta-analysis in genome-wide association studies. *Pharmacogenomics* 10 (2): 191-201.

57. Wood A.R., Esko T., Yang J., Vedantam S., Pers T.H., et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* 46 (11): 1173-1186.

58. Macarthur J., Bowler E., Cerezo M., Gil L., Hall P., et al. (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research* 45 (D1): D896-D901.

59. Koonin E.V., Galperin M.Y. (2003) Chapter 5: Genome Annotation and Analysis. *Sequence - Evolution - Function: Computational Approaches in Comparative Genomics*. Boston: Kluwer Academic.

60. Hrdlickova B., De Almeida R.C., Borek Z., Withoff S. (2014) Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1842 (10): 1910-1922.

61. Khurana E., Fu Y., Chakravarty D., Demichelis F., Rubin M.A., et al. (2016) Role of non-coding sequence variants in cancer. *Nature Reviews Genetics* 17 (2): 93-108.

62. Kumar V., Westra H.J., Karjalainen J., Zhernakova D.V., Esko T., et al. (2013) Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLOS Genetics* 9 (1): e1003201.

63. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489 (7414): 57-74.

64. ENCODE Project Consortium (2017) Encyclopedia of DNA elements at UCSC. URL: https://genome.ucsc.edu/ENCODE/. Accssed on 2017-04-22, 2017

65. Laber S., Cox R.D. (2017) Mouse models of human GWAS hits for obesity and diabetes in the post genomic era: Time for reevaluation. *Frontiers in Endocrinology* 8: 11.

66. Cao C., Moult J. (2014) GWAS and drug targets. *BMC Genomics* 15 (4).

67. Zhang J., Jiang K., Lv L., Wang H., Shen Z., et al. (2015) Use of genome-wide association studies for cancer research and drug repositioning. *PLOS One* 10 (3): e0116477.

68. Abraham G., Inouye M. (2015) Genomic risk prediction of complex human disease and its clinical application. *Current Opinion in Genetics and Development* 33: 10-16.

69. Eschrich S., Yang I., Bloom G., Kwong K.Y., Boulware D., et al. (2005) Molecular staging for survival prediction of colorectal cancer patients. *Journal of Clinical Oncology* 23 (15): 3526-3535.

70. Lall K., Magi R., Morris A., Metspalu A., Fischer K. (2017) Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. *Genetics in Medicine* 19 (3): 322-329.

71. Manor O., Segal E. (2013) Predicting disease risk using bootstrap ranking and classification algorithms. *PLOS Computational Biology* 9 (8): e1003200.

72. Vazquez A.I., Veturi Y., Behring M., Shrestha S., Kirst M., et al. (2016) Increased proportion of variance explained and prediction accuracy of survival of breast cancer patients with use of whole-genome multiomic profiles. *Genetics* 203 (3): 1425-1438.

73. Silventoinen K., Sammalisto S., Perola M., Boomsma D.I., Cornes B.K., et al. (2003) Heritability of adult body height: A comparative study of twin cohorts in eight countries. *Twin Research and Human Genetics* 6 (5): 399-408.

74. Mataix-Cols D., Boman M., Monzani B., Rück C., Serlachius E., et al. (2013) Population-based, multigenerational family clustering study of obsessive-compulsive disorder. *JAMA Psychiatry* 70 (7): 709-717.

75. Zuk O., Hechter E., Sunyaev S.R., Lander E.S. (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* 109 (4): 1193-1198.

76. Zaitlen N., Kraft P., Patterson N., Pasaniuc B., Bhatia G., et al. (2013) Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLOS Genetics* 9 (5): e1003520.

77. Gibson G. (2012) Rare and common variants: twenty arguments. *Nature Reviews Genetics* 13 (2): 135-145.

78. Wray N.R., Yang J., Hayes B.J., Price A.L., Goddard M.E., et al. (2013) Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics* 14 (7): 507-515.

79. Mucci L.A., Hjelmborg J.B., Harris J.R., Czene K., Havelick D.J., et al. (2016) Familial risk and heritability of cancer among twins in Nordic countries. *JAMA* 315 (1): 68-76.

80. Visscher P.M., Andrew T., Nyholt D.R. (2008) Genome-wide association studies of quantitative traits with related individuals: little (power) lost but much to be gained. *European Journal of Human Genetics* 16: 387-390.
81. Tenesa A., Haley C.S. (2013) The heritability of human disease: estimation, uses and abuses. *Nature Reviews Genetics* 14 (2): 139-149.
82. Fisher R.A. (1952) Statistical methods in genetics. *Heredity* 6 (1): 1-12.
83. von Hinke S., Davey S.G., Lawlor D.A., Propper C., Windmeijer F. (2016) Genetic markers as instrumental variables. *Journal of Health Economics* 45: 131-148.
84. Buchanan J.A., Scherer S.W. (2008) Contemplating effects of genomic structural variation. *Genetics in Medicine* 10 (9): 639-647.
85. Belton J.M., Mccord R.P., Gibcus J.H., Naumova N., Zhan Y., et al. (2012) Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58 (3): 268-276.
86. Lappalainen T., Sammeth M., Friedländer M.R., 'T Hoen P.A.C., Monlong J., et al. (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501 (7468): 506-511.
87. Martincorena I., Campbell P.J. (2015) Somatic mutation in cancer and normal cells. *Science* 349 (6255): 1483-1489.
88. Watson I.R., Takahashi K., Futreal P.A., Chin L. (2013) Emerging patterns of somatic mutations in cancer. *Nature Reviews Genetics* 14: 703.
89. Luzzatto L. (2012) Sickle cell anaemia and malaria. *Mediterranean Journal of Hematology and Infectious Diseases* 4 (1): e2012065.
90. Charlesworth D., Willis J.H. (2009) The genetics of inbreeding depression. *Nature Reviews Genetics* 10 (11): 783-796.
91. Robinson M.R., Kleinman A., Graff M., Vinkhuyzen A.A.E., Couper D., et al. (2017) Genetic evidence of assortative mating in humans. *Nature Human Behaviour* 1: 0016.
92. Silventoinen K., Kaprio J., Lahelma E., Viken R.J., Rose R.J. (2003) Assortative mating by body height and BMI: Finnish twins and their spouses. *American Journal of Human Biology* 15 (5): 620-627.
93. Falconer D.S., Mackay T.F.C. (1996) *Introduction to Quantitative Genetics*. 4 ed. Essex: Pearson Education Limited.

94. Lynch M., Walsh B. (1998) *Genetics and Analysis of Quantitative Traits*. 1 ed. Sunderland, MA: Sinauer Associates, Inc.

95. Sutker P., Tabakoff B., Goist K.J., Randall C.L. (1983) Acute alcohol intoxication, mood states and alcohol metabolism in women and men. *Pharmacology Biochemistry and Behavior* 18 (Suppl 1): 349-354.

96. Correia C., Oliveira G., Vicente A.M. (2014) Protein interaction networks reveal novel autism risk genes within GWAS statistical noise. *PLOS One* 9 (11): e112399.

97. Kar S.P., Tyrer J.P., Li Q., Lawrenson K., Aben K.K., et al. (2015) Network-based integration of GWAS and gene expression identifies a HOX-centric network associated with serous ovarian cancer risk. *Cancer Epidemiology, Biomarkers and Prevention* 24 (10): 1574-1584.

98. Breen G., Li Q., Roth B.L., O'donnell P., Didriksen M., et al. (2016) Translating genome-wide association findings into new therapeutics for psychiatry. *Nature Neuroscience* 19 (11): 1392-1396.

99. Shah S., Bonder M.J., Marioni R.E., Zhu Z., Mcrae A.F., et al. (2015) Improving phenotypic prediction by combining genetic and epigenetic associations. *American Journal of Human Genetics* 97 (1): 75-85.

100. Xia C., Amador C., Huffman J., Trochet H., Campbell A., et al. (2016) Pedigree- and SNP-associated genetics and recent environment are the major contributors to anthropometric and cardiometabolic trait variation. *PLOS Genetics* 12 (2): e1005804.

101. Wray N.R., Goddard M.E., Visscher P.M. (2007) Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research* 17 (10): 1520-1528.

102. Fisher R.A. (1919) XV.—The correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* 52 (02): 399-433.

103. Lango Allen H., Estrada K., Lettre G., Berndt S.I., Weedon M.N., et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467 (7317): 832-838.

104. Teslovich T.M., Musunuru K., Smith A.V., Edmondson A.C., Stylianou I.M., et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466 (7307): 707-713.

105. Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorff L.A., et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461 (7265): 747-753.

106. Benjamin D.J., Cesarini D., van der Loos M.J.H.M., Dawes C.T., Koellinger P.D., et al. (2012) The genetic architecture of economic and political preferences. *Proceedings of the National Academy of Sciences* 109 (21): 8026-8031.

107. Vattikuti S., Guo J., Chow C.C. (2012) Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLOS Genetics* 8 (3): e1002637.

108. Krakauer N.Y., Krakauer J.C. (2012) A new body shape index predicts mortality hazard independently of body mass index. *PLOS One* 7 (7): e39504.

109. Keller M.C., Garver-Apgar C.E., Wright M.J., Martin N.G., Corley R.P., et al. (2013) The genetic correlation between height and IQ: shared genes or assortative mating? *PLOS Genetics* 9 (4): e1003451.

110. Silventoinen K., Magnusson P.K., Tynelius P., Kaprio J., Rasmussen F. (2008) Heritability of body size and muscle strength in young adulthood: a study of one million Swedish men. *Genetic Epidemiology* 32 (4): 341-349.

111. Schousboe K., Visscher P.M., Erbas B., Kyvik K.O., Hopper J.L., et al. (2004) Twin study of genetic and environmental influences on adult body size, shape, and composition. *International Journal of Obesity and Related Metabolic Disorders* 28 (1): 39-48.

112. Goode E.L., Cherny S.S., Christian J.C., Jarvik G.P., de Andrade M. (2007) Heritability of longitudinal measures of body mass index and lipid and lipoprotein levels in aging twins. *Twin Research and Human Genetics* 10 (05): 703-711.

113. Selby J.V., Newman B., Quesenberry C.P., Jr., Fabsitz R.R., Carmelli D., et al. (1990) Genetic and behavioral influences on body fat distribution. *International Journal of Obesity* 14 (7): 593-602.

114. Hunter D.J., Lange M., Snieder H., MacGregor A.J., Swaminathan R., et al. (2002) Genetic contribution to renal function and electrolyte balance: a twin study. *Clinical Science* 103 (3): 259-265.

115. Bathum L., Fagnani C., Christiansen L., Christensen K. (2004) Heritability of biochemical kidney markers and relation to survival in the elderly—results from a Danish population-based twin study. *Clinica Chimica Acta: International Journal of Clinical Chemistry* 349 (1–2): 143-150.

116. Katoh S., Lehtovirta M., Kaprio J., Harjutsalo V., Koskenvuo M., et al. (2005) Genetic and environmental effects on fasting and postchallenge plasma glucose and serum insulin values in Finnish twins. *The Journal of Clinical Endocrinology and Metabolism* 90 (5): 2642-2647.

117. Snieder H., Harshfield G.A., Treiber F.A. (2003) Heritability of Blood Pressure and Hemodynamics in African- and European-American Youth. *Hypertension* 41 (6): 1196-1201.

118. Smith B.H., Campbell A., Linksted P., Fitzpatrick B., Jackson C., et al. (2012) Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *International Journal of Epidemiology* 42 (3): 689-700.

119. Amador C., Huffman J., Trochet H., Campbell A., Porteous D., et al. (2015) Recent genomic heritage in Scotland. *BMC Genomics* 16 (1): 437.

120. Scottish Goverment (2014) Scottish index of multiple deprivation. URL: http://www.scotland.gov.uk/Topics/Statistics/SIMD. Accssed on 12-Sep, 2014

121. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81 (3): 559-575.

122. Aulchenko Y.S., Ripke S., Isaacs A., van Duijn C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23 (10): 1294-1296.

123. R Development Core Team (2010) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

124. Fisher R.A. (1930) *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

125. Yang J., Bakshi A., Zhu Z., Hemani G., Vinkhuyzen A.A.E., et al. (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics* 47 (10): 1114-1120.

126. Jensen A.R. (1998) *The g Factor: The Science of Mental Ability.* Westport, CT: Praeger.

127. Zeng Y., Navarro P., Xia C., Amador C., Fernandez-Pujals A.M., et al. (2016) Shared genetics and couple-associated environment are major contributors to the risk of both clinical and self-declared depression. *EBioMedicine* 14: 161-167.

128. Hill W.D., Arslan R.C., Xia C., Luciano M., Amador C., et al. (2018) Genomic analysis of family data reveals additional genetic effects on intelligence and personality. *Molecular Psychiatry* (ePrint).

129. Mathews C.A., Reus V.I. (2001) Assortative mating in the affective disorders: a systematic review and meta-analysis. *Comprehensive Psychiatry* 42 (4): 257-262.

130. Farley F., Davis S.A. (1997) Arousal, personality, and assortative mating in marriage. *Journal of Sex and Marital Therapy* 3 (2): 122-127.

131. Eaves L.J. (1973) Assortative mating and intelligence: An analysis of pedigree data. *Heredity* 30 (2): 199-210.

132. Amador C., Xia C., Nagy R., Campbell A., Porteous D., et al. (2017) Regional variation in health is predominantly driven by lifestyle rather than genetics. *Nature Communications* 8 (1): 801.

133. Davies G., Tenesa A., Payton A., Yang J., Harris S.E., et al. (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry* 16 (10): 996-1005.

134. Kirkpatrick R.M., McGue M., Iacono W.G., Miller M.B., Basu S. (2014) Results of a “GWAS Plus:” General Cognitive Ability Is Substantially Heritable and Massively Polygenic. *PLOS One* 9 (11): e112390.

135. Posthuma D., De Geus E.J.C., Boomsma D.I. (2001) Perceptual speed and IQ are associated through common genetic factors. *Behavior Genetics* 31 (6): 593-602.

136. Smith D.J., Escott-Price V., Davies G., Bailey M.E., Colodro-Conde L., et al. (2016) Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Molecular psychiatry* 21 (11): 1644.

137. van den Berg S.M., de Moor M.H., Verweij K.J., Krueger R.F., Luciano M., et al. (2016) Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behavior Genetics* 46 (2): 170-182.

138. Lo M.T., Hinds D.A., Tung J.Y., Franz C., Fan C.C., et al. (2017) Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nature Genetics* 49 (1): 152-156.

139. Vukasović T., Bratko D. (2015) Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin* 141 (4): 769.

140. Lubke G.H., Hottenga J.J., Walters R., Laurin C., De Geus E.J.C., et al. (2012) Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biological Psychiatry* 72 (8): 707-709.

141. Cross-Disorder Group Of The Psychiatric Genomics Consortium (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* 45 (9): 984-994.

142. Sullivan P.F., Neale M.C., Kendler K.S. (2000) Genetic epidemiology of major depression: Review and meta-analysis. *The American Journal of Psychiatry* 157 (10): 1552-1562.

143. Canela-Xandri O., Law A., Gray A., Woolliams J.A., Tenesa A. (2015) A new tool called DISSECT for analysing large genomic data sets using a big data approach. *Nature Communications* 6: 10162.

144. Spielman R.S., Mcginnis R.E., Ewenst W.J. (1993) Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics* 52: 506-516.

145. Lake S.L., Blacker D., Laird N.M. (2000) Family-based tests of association in the presence of linkage. *American Journal of Human Genetics* 67: 1515-1525.

146. Martin E.R., Bass M.P., Hauser E.R., Kaplan N.L. (2003) Accounting for linkage in family-based tests of association with missing parental genotypes. *American Journal of Human Genetics* 73 (5): 1016-1026.

147. Freedman M.L., Reich D., Penney K.L., Mcdonald G.J., Mignault A.A., et al. (2004) Assessing the impact of population stratification on genetic association studies. *Nature Genetics* 36 (4): 388-393.

148. Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A., et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38 (8): 904-909.

149. Yu J., Pressoir G., Briggs W.H., Vroh Bi I., Yamasaki M., et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38 (2): 203-208.

150. Aulchenko Y.S., De Koning D.J., Haley C. (2007) Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* 177 (1): 577-585.

151. Chen W.M., Abecasis G.R. (2007) Family-based association tests for genomewide association scans. *American Journal of Human Genetics* 81 (5): 913-926.

152. Yang J., Zaitlen N.A., Goddard M.E., Visscher P.M., Price A.L. (2014) Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics* 46 (2): 100-106.

153. Burdett T., Hall P.N., Hastings E., Hindorff L.A., Junkins H.A., et al. (2017) The NHGRI-EBI Catalog of published genome-wide association studies. URL: www.ebi.ac.uk/gwas. Accssed on Janurary, 2017

154. Johnson A.D., Handsaker, R. E., Pulit, S., Nizzari, M. M., O'Donnell, C. J., de Bakker, P. I. W. (2008) SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24 (24): 2938-2939.

155. Grant S.F.A., Bradfield J.P., Zhang H., Wang K., Kim C.E., et al. (2009) Investigation of the locus near MC4R with childhood obesity in Americans of European and African ancestry. *Obesity (Silver Spring)* 17 (7): 1461-1465.

156. Wheeler E., Huang N., Bochukova E.G., Keogh J.M., Lindsay S., et al. (2013) Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nature Genetics* 45 (5): 513-517.

157. Taylor A.E., Sandeep M.N., Janipalli C.S., Giambartolomei C., Evans D.M., et al. (2011) Associations of FTO and MC4R variants with obesity traits in Indians and the role of rural/urban environment as a possible effect modifier. *Journal of Obesity* 2011: 307542.

158. Vasan S.K., Fall T., Neville M.J., Antonisamy B., Fall C.H., et al. (2012) Associations of variants in FTO and near MC4R with obesity traits in South Asian Indians. *Obesity (Silver Spring)* 20 (11): 2268-2277.

159. Wang T., Ma X., Peng D., Zhang R., Sun X., et al. (2016) Effects of obesity related genetic variations on visceral and subcutaneous fat distribution in a Chinese population. *Scientific Reports* 6: 20691.

160. Wu M., Michaud E.J., Johnson D.K. (2003) Cloning, functional study and comparative mapping of Luzp2 to mouse chromosome 7 and human chromosome 11p13-11p14. *Mammalian Genome* 14 (5): 323-334.

161. Han J.C., Liu Q.R., Jones M., Levinn R.L., Menzie C.M., et al. (2008) Brain-derived neurotrophic factor and obesity in the WAGR syndrome. *The New England Journal of Medicine* 359 (9): 918-927.

162. Rodriguez-Lopez R., Perez J.M., Balsera A.M., Rodriguez G.G., Moreno T.H., et al. (2013) The modifier effect of the BDNF gene in the phenotype of the WAGRO syndrome. *Gene* 516 (2): 285-290.

163. van Rooij F.J., Qayyum R., Smith A.V., Zhou Y., Trompet S., et al. (2017) Genome-wide trans-ethnic meta-analysis identifies seven genetic loci influencing erythrocyte traits and a role for RBPMS in erythropoiesis. *American Journal of Human Genetics* 100 (1): 51-63.

164. Kottgen A., Pattaro C., Boger C.A., Fuchsberger C., Olden M., et al. (2010) New loci associated with kidney function and chronic kidney disease. *Nature Genetics* 42 (5): 376-384.

165. Nagy R., Boutin T.S., Marten J., Huffman J.E., Kerr S.M., et al. (2017) Exploration of haplotype research consortium imputation for genome-wide

association studies in 20,032 Generation Scotland participants. *Genome Medicine* 9 (1): 23.

166. Tenesa A., Rawlik K., Navarro P., Canela-Xandri O. (2016) Genetic determination of height-mediated mate choice. *Genome Biology* 16 (1): 269.

167. Damen J.A., Hooft L., Schuit E., Debray T.P., Collins G.S., et al. (2016) Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 353: i2416.

168. Dite G.S., Macinnis R.J., Bickerstaffe A., Dowty J.G., Allman R., et al. (2016) Breast cancer risk prediction using clinical models and 77 independent risk-associated SNPs for women aged Under 50 Years: Australian breast cancer family registry. *Cancer Epidemiology, Biomarkers and Prevention* 25 (2): 359-365.

169. Henderson C.R. (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31 (2): 423-447.

170. Morota G., Gianola D. (2014) Kernel-based whole-genome prediction of complex traits: a review. *Frontiers in Genetics* 5: 363.

171. Gianola D., Fernando R.L., Stella A. (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173 (3): 1761-1776.

172. Ober U., Erbe M., Long N., Porcu E., Schlather M., et al. (2011) Predicting genetic values: a kernel-based best linear unbiased prediction with genomic data. *Genetics* 188 (3): 695-708.

173. Tikhonov A.N., Arsenin V.Y. (1977) *Solutions of Ill-Posed Problems*; F. John TE, editor. Washington: V. H. Winston & Sons.

174. Tikhonov A.N., Goncharsky A., Stepanov V.V., Yagola A.G. (1995) *Numerical Methods for the Solution of Ill-Posed Problems*. The Netherlands: Springer Netherlands.

175. Aronszajn N. (1950) Theory of reproducing kernels. *Transactions of the American Mathematical Society* 68 (3): 337-404.

176. Hofmann M. (2016) Support vector machines -- kernels and the kernel trick. URL: http://www.cogsys.wiai.uni-bamberg.de/teaching/ss06/hs_svm/slides/SVM_Seminarbericht_Hofmann.pdf. Accssed on 08/09, 2016

177. Gianola D., Van Kaam J.B. (2008) Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178 (4): 2289-2303.

178. Kimeldorf G., Wahba, G. (1971) Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications* 33: 82-95.

179. Lanckriet G.R.G., Cristianini N., Bartlett P., Ghaoui L.E., Jordan M.I. (2004) Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research* 5: 27-72.

180. Musunuru K., Ingelsson E., Fornage M., Liu P., Murphy A.M., et al. (2017) The expressed genome in cardiovascular diseases and stroke: Refinement, diagnosis, and prediction: A scientific statement from the American Heart Association. *Circulation Cardiovascular Genetics* 10 (4): e000037.

181. Spiliopoulou A., Nagy R., Bermingham M.L., Huffman J.E., Hayward C., et al. (2015) Genomic prediction of complex human traits: relatedness, trait architecture and predictive meta-models. *Human Molecular Genetics* 24 (14): 4167-4182.

182. Makowsky R., Pajewski N.M., Klimentidis Y.C., Vazquez A.I., Duarte C.W., et al. (2011) Beyond missing heritability: Prediction of complex traits. *PLOS Genetics* 7 (4): e1002051.

183. Garrod A. (1908) Inborn errors of metabolism. *The Lancet* 2: 1-7.

184. Garrod A. (1902) The incidence of alkaptonuria: A study in chemical individuality. *The Lancet* 2: 1616-1620.

185. Voight B.F., Kudaravalli S., Wen X., Pritchard J.K. (2006) A map of recent positive selection in the human genome. *PLOS Biology* 4 (3): e72.

186. Keinan A., Mullikin J.C., Patterson N., Reich D. (2007) Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genetics* 39 (10): 1251.

187. Wikipedia Contributors (2017) Assortative mating. URL: https://en.wikipedia.org/wiki/Assortative_mating. Accssed on Oct 2017, 2017

188. Lynch M., Walsh B. (1998) Chapter 7: Resemblance Between Relatives. *Genetics and Analysis of Quantitative Traits*. 1 ed. Sunderland, MA: Sinauer Associates, Inc.

189. Schousboe K., Willemsen G., Kyvik K.O., Mortensen J., Boomsma D.I., et al. (2012) Sex differences in heritability of BMI: A comparative study of results from Twin studies in eight countries. *Twin Research* 6 (5): 409-421.

190. Sullivan P.F., Kendler K.S., Neale M.C. (2003) Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Archives of General Psychiatry* 60 (12): 1187-1192.

191. Magnusson P.K., 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. *International Journal of Obesity and Related Metabolic Disorders* 26 (9): 1225-1231.

192. Stulp G., Simons M.J.P., Grasman S., Pollet T.V. (2017) Assortative mating for human height: A meta-analysis. *American Journal of Human Biology* 29 (1): e22917.

193. Allison D.B., Neale M.C., Kezis M.I., Alfonso V.C., Heshka S., et al. (1996) Assortative mating for relative weight: Genetic implications. *Behavior Genetics* 26 (2): 103-111.

194. Johansson K., Neovius M., Hemmingsson E. (2014) Effects of anti-obesity drugs, diet, and exercise on weight-loss maintenance after a very-low-calorie diet or low-calorie diet: a systematic review and meta-analysis of randomized controlled trials. *The American Journal of Clinical Nutrition* 99 (1): 14-23.

195. Kerksick C.M., Wismann-Bunn J., Fogt D., Thomas A.R., Taylor L., et al. (2010) Changes in weight loss, body composition and cardiovascular disease risk after altering macronutrient distributions during a regular exercise program in obese women. *Nutrition Journal* 9: 59-59.

196. Falconer D.S., Mackay T.F.C. (1996) Chapter 10: Heritability. *Introduction to Quantitative Genetics*. 4 ed. Essex: Pearson Education Limited.

Supplementary Materials

SP Texts

Text S2.1

Text S2.1 Simulating phenotypes

In order to evaluate the robustness of our models and the performance of our stepwise model selection, we conducted a simulation study. We simulated, based on the real genotypic information and the real pedigree, different sets of phenotypes for each of the 9,863 individuals in GS10K. The simulated phenotypes were generated by combining various proportions of simulated effects for SNP-associated genetics, pedigree-associated genetics, nuclear family environment, shared couple environment and sibling environment.

For simulating the genetic effects we used a similar approach to Zaitlen *et al.* [1], but based on our real genotype information in GS10K. The genome was divided into two: even chromosomes were used to create the observed genetic effects that were in LD with the SNPs (in a later step, only even chromosomes will be used to generate the appropriate genomic relationship matrices); odd chromosomes were used to create the unobserved variants that were not in LD with the SNP array. We randomly selected 1 in every 500 SNPs (MAF > 0.05) on even chromosomes, ending up with 550 'causal loci' representing the causal variants tagged by genotyping platform. We assigned an effect to the rare alleles of the selected markers (assuming an additive model). The summed effect for those loci was $\mathbf{g_g}$ and was calculated as $\sum_{i=1}^{N} a_i x_i$ for each individual, where *N* is the number of causal loci, a_i is the effect size of allele *i* and x_i is the allelic dose for allele *i*. Similarly, another 550 common 'causal loci' were randomly selected on odd chromosomes, representing the variants that were not in LD with the SNP array. The summed effect for those loci was $\mathbf{g_{kin}}$, which was calculated using the same formula as $\mathbf{g_g}$.

These basic genetic settings were the same as in Zaitlen *et al.* [1], except for the assumption we used for effect sizes. The effect sizes in our simulation study were derived from an exponential distribution as in Fisher [2]. The distributions of effect

sizes for 'casual loci' on even and odd chromosomes were $\boldsymbol{a_g} \sim E(\lambda = \sqrt{\frac{4\times N\times \overline{pq}}{h_g^2}})$ and $\boldsymbol{a_{kin}} \sim E(\lambda = \sqrt{\frac{4\times N\times \overline{pq}}{h_{kin}^2}})$ respectively, where $\boldsymbol{a_g}$ and $\boldsymbol{a_{kin}}$ are $N \times 1$ vectors of effect sizes for chosen SNPs on even and odd chromosomes separately and $\overline{pq}$ is the mean of minor allele frequency times major allele frequency for these loci, which is 0.1825 here.

We transformed $\mathbf{g_g}$ and $\mathbf{g_{kin}}$ to normal distributions with mean equal to 0 and variance equal to h_g^2 or h_{kin}^2, which are the proportion of the variance of the simulated phenotypes explained by variants in LD with the markers, and variants not in LD with the markers, respectively.

The environmental effects were simulated based on the real pedigree. For sibling environment ($\mathbf{e_s}$) and couple environment ($\mathbf{e_c}$), the effect sizes were derived from two normal distributions: $N(0, e_s^2)$ and $N(0, e_c^2)$ respectively. We assigned the same random couple effect to each pair of individuals in a couple and the same random sibling effect to each of the full-siblings from the same nuclear family. Individuals without any spouse or/and siblings in the data were also given a random couple effect and a random sibling effect that was unique to themselves. For the nuclear family environment ($\mathbf{e_f}$), individuals were given two nuclear family effects: one for their youth (representing familial environment when living with their parents) and the other for adulthood (familial environment when living alone or with their spouse and children). Nuclear family members shared the same nuclear family effect, whereas single individuals did not share nuclear family effect with any other individuals. Therefore, individuals with parents and with a spouse or/and offspring (1,305 individuals in GS10K) shared two separate familial effects, one with their parents and any sibs and one with their spouse and/or children; individuals without any first degree relatives (1,785 individuals in GS10K) had two unique familial effects; and the remaining individuals (6,773 individuals in GS10K) had one shared and one unique familial effects. Both family environments (youth and adulthood) contributed equally to create the final family environmental effect. The rationale for this approach is that in three generation families (grandparents, parents and progeny) it allows separate nuclear family environment effects for the grandparents and their grandprogeny, i.e. it

does not assume that they share the same family environment. The effect size was randomly drawn from $N(0, e_f^2)$.

As before, we transformed $\mathbf{e_f}$, $\mathbf{e_s}$ and $\mathbf{e_c}$ to normal distributions with mean equal to 0 and variance equal to e_f^2, e_s^2 or e_c^2, being e_f^2, e_s^2 or e_c^2 the proportion of the variance of the simulated phenotype explained by family, sibling or couple environment, respectively.

We also simulated a random residual effect for each individual ($\boldsymbol{\varepsilon}$), the residuals were derived from $N(0, e_e^2)$ where e_e^2 represents the proportion of variance remaining in each of the scenarios. For each scenario, each component ($h_g^2, h_{kin}^2, e_c^2, e_s^2, e_f^2$) was given a proportion of the variance explained (0% - 50%) and e_e^2 was $1 - h_g^2 - h_{kin}^2 - e_c^2 - e_s^2 - e_f^2$. The final phenotypes would be the sum of transformed $\mathbf{g_g}$, $\mathbf{g_{kin}}$, $\mathbf{e_f}$, $\mathbf{e_s}$, $\mathbf{e_c}$ and $\boldsymbol{\varepsilon}$, and the expected mean and variance of simulated phenotypes was 0 and 1, respectively.

1. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, et al. (2013) Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLOS Genetics* 9: e1003520.
2. Fisher RA (1930) *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Text S4.1

Text S4.1 The expected number of individuals benefits from modelling sibling environment, family environment and pedigree-associated genetics in the extended GWAS method

By counting the elements in $\mathbf{GRM_{kin}}$ matrix, I found that, for individuals with relatives (sharing pedigree-associated genetic effects) in GS10K, the average number of relatives they have is 2.78 and I take the nearest integer of 3. The probability that 4 individuals (3 relatives + the individual himself/herself) are all in the training set is 0.8^4=0.4096 (Situation I), all in the validation set is 0.2^4=0.0016 (Situation II), 1 in the training set and the rest in the validation set is $4*0.2*0.8^3$=0.4096 (Situation III), half in the training set and half in the validation set is $6*0.2^2*0.8^2$=0.1536 (Situation IV) and 3 in the training set and the rest in the validation set is $4*0.2^3*0.8$=0.0256 (Situation V). The possibilities sum to 1.

The unique non-zero elements for $\mathbf{GRM_{kin}}$ matrix is 8080 pairs. The number of pairwise relationship for a group of 4 individuals is 4*3/2=6. 8080/6=1346.67. Therefore, I assume that there are 1347 different groups of individuals (e.g. an extended family), 4 individuals per group, and individuals within each group sharing pedigree-associated genetic effects but not between groups.

If a group is in Situation I or II, then none of the group member benefits from modelling pedigree-associated genetic effects. If a group is in Situation III, IV or V, then all of them benefit from modelling pedigree-associated genetic effects. Thus, the expected number of individuals benefited from modelling pedigree-associated effects is $4*1347*(0.4096+0.1536+0.0256) \approx 3172$. Note, the number of unrelated individuals in GS10K is ~6k, i.e. there are 4k individuals who have relatives in the data. This means that over three fourths of the people who have blood relatives in GS10K or one third of the whole population might benefit from modelling pedigree-associated genetic effects.

The average number of siblings for individuals who have siblings in GS10K (sharing common sibling environment) is 1.22 and I take the nearest integer of 1. The probability that 2 individuals (1 siblings + the individual himself/herself) are both in the training set is 0.8^2=0.64 (Situation I), both in the validation set is 0.2^2=0.04

(Situation II) and 1 in the training set and 1 in the validation set is 2*0.2*0.8=0.32 (Situation III). The possibilities sum to 1.

The unique non-zero elements for $\mathbf{ERM_{Sib}}$ matrix is 676 pairs. The number of pairwise relationship for a group of 2 individuals is 2*1/2=1. Therefore, 676 pairs of non-zero elements in S matrix could be considered as there are 676 groups of sibling pairs from 676 different families, i.e. each pair of sibling share sibling environment within group but not between groups.

If a sibling pair is in Situation I or II, then neither of them benefits from modelling sibling environment. If a sibling pair is in Situation III, then both of them benefit from modelling sibling environment. Thus, the expected number of individuals benefited from sibling environment is $2*676*0.32 \approx 432$ (~200 pairs of siblings). Therefore, one third of the sibling pairs or ~4% of the whole population might benefit from modelling sibling environment.

The average number of nuclear family members for individuals who have nuclear family members in GS10K (sharing common family environment) is 2.25 and I take the nearest integer of 2. The probability that 3 individuals (2 family members + the individual himself) are all in the training set is 0.8^3=0.512 (Situation I) and all in the validation set is 0.2^3=0.008 (Situation II).

The unique non-zero elements for F matrix is 4821 pairs. The number of pairwise relationship for a group of 3 individuals is 3*2/2=3. Therefore, roughly there are 4821/3= 1607 nuclear families in GS10K, 3 individuals per family.

If a nuclear family is in Situation I or II, then none of the nuclear family members benefits from modelling family environment. If a nuclear family is not in Situation I and II, then all the family members benefit from modelling family environment. Thus, the expected number of individuals benefit from modelling family environment is $3*1607*(1-0.512-0.08) \approx 2314$. Approximately half of the nuclear families or a quarter of the whole population might benefit from modelling family environment.

Text S5.1

Text S5.1 Estimation of standard errors in sib-sib and sib-in-law pedigree design

Delta method is a method to estimate the standard errors (S.E.) of a ratio and it was used to estimate the S.E. of ρ and h^2_{AM} in sib-sib and sib-in-law pedigree study in Chapter 5. The code used in R for delta method is as follows.

```
mr <- function(x, sigmax, y, sigmay) {
  eff <- sum(x*y/sigmay^2, na.rm = TRUE)/sum(x^2/sigmay^2, na.rm = TRUE)
  err <- sqrt(1/sum(x^2/sigmay^2, na.rm = TRUE))
  p <- pchisq(eff**2/err**2, 1, lower.tail = FALSE)
  return(list(eff = eff, err = err, p = p))
}
```

Therefore, by using this mr function in R, it is possible to obtain the S.E. of y/x (se(y/x)), which equals mr(x , se(x) , y , se(y)).

Since ρ was estimated as the ratio of r_{FSIL} and r_{FS} ($\rho = \frac{r_{FSIL}}{r_{FS}}$), the S.E. of ρ (se(ρ)) equals mr(r_{FS} , se(r_{FS}) , r_{FSIL} , se(r_{FSIL})).

Since h^2_{AM} was estimated as the ratio of $-1+\sqrt{1+8r_{FSIL}}$ and 2ρ ($h^2_{AM} = \frac{-1+\sqrt{1+8r_{FSIL}}}{2\rho}$), to estimate the S.E. of h^2_{AM} we need to know the S.E. of the numerator and denominator first.

The S.E of 2ρ is twice the S.E. of ρ, i.e. se(2ρ)=2se(ρ)

The S.E. of $-1+\sqrt{1+8r_{FSIL}}$ is the same as the S.E. of $\sqrt{8r_{FSIL}}$. The following procedure is how I get a generalised equation which links se(X) to se($\sqrt{X}$) to estimate se($\sqrt{8r_{FSIL}}$).

For $X \sim N(0, \sigma^2)$, $Var(X^2) = 2[Var(X)]^2$, see https://math.stackexchange.com/questions/620045/mean-and-variance-of-squared-gaussian-y-x2-where-x-sim-mathcaln0-sigma (accessed, 1-Dec, 2017) and

http://math.arizona.edu/~jwatkins/H_expectedvalue.pdf , page 134 (accessed, 1-Dec, 2017).

Besides, $se(X) = \frac{Var(X)}{\sqrt{n}}$.

Therefore, $Var(X^2) = se(X^2)\sqrt{n} = 2[se(X)\sqrt{n}]^2$

$$se(X) = \sqrt{\frac{se(X^2)}{2\sqrt{n}}}$$

By replacing X with $\sqrt{8r_{FSIL}}$, it is possible to obtain the S.E. of the numerator.

$$se\left(-1 + \sqrt{1 + 8r_{FSIL}}\right) = se\left(\sqrt{8r_{FSIL}}\right) = \sqrt{\frac{4se(r_{FSIL})}{\sqrt{n}}}$$

After I got $se\left(-1 + \sqrt{1 + 8r_{FSIL}}\right)$ and $se(2\rho)$, I put them into mr function in R to estimate the S.E. of h^2_{AM}.

$se(h^2_{AM})$=mr(2ρ, $2se(\rho)$, $-1 + \sqrt{1 + 8r_{FSIL}}$, $\sqrt{\frac{4se(r_{FSIL})}{\sqrt{n}}}$).

Thereby, it is possible to estimate the S.E. of ρ and h^2_{AM} by using the observed phenotypic correlations between sib-sib and between sib-sib-in-law and the S.E. of those observations.

However, $se(X) = \sqrt{\frac{se(X^2)}{2\sqrt{n}}}$ is true when $X \sim N(0, \sigma^2)$ and the distribution of $\sqrt{8r_{FSIL}}$ apparently is different than that (mean is not 0 and has a limited interval $[0, 2\sqrt{2}]$), which might cause bias in the estimation of $se(h^2_{AM})$.

SP Tables

Table S2.1

Table S2.1 Descriptive analysis of traits and covariates using GS10K data.

Name	Category	
Height	Traits	Body Measurements
Weight		
Fat		
BMI		
Hips		
Waist		
WHR		
ABSI		
Urea		Biochemistry[a,b,c]
Creatinine		
Glucose		
TC		
HDL		
SBP		Blood Pressure & Heart Rate[d]
DBP		
HR		
Age	Covariates	
Sex		
SIMD		
Centre		
20PCs		

[a.] GS:SFHS blood samples were analysed using standard automated methods in the NHS biochemistry laboratories local to the clinics across Scotland in which recruitment took place. These laboratories participate in the UK National External Quality Assessment Service (UKNEQAS) scheme and all tests had inter-assay CVs <5%.

[b.] Measurements of TC, HDL, Creatinine and Urea were from serum prepared from 5ml of venous blood collected into a tube containing clot activator & gel separator.

[c.] For glucose measurement, 2ml of venous blood was taken from consenting participants in GS research clinics using standard venepuncture procedures and collected in a sodium fluoride / potassium oxalate tube, with fasting duration recorded. The Glasgow lab used the Abbott Architect, hexokinase/glucose-6-phosphate dehydrogenase method for measurement of glucose.

[d.] Blood Pressure & Heart Rate traits were measured twice and here we use the 2nd measurement.

Name	Description	Primary Data
Height	Body Height	Yes
Weight	Body Weight	Yes
Fat	Body Fat Composition (Tanita scales)	Yes
BMI	$Weight/Height^2$	No
Hips	Hip Circumference	Yes
Waist	Waist Circumference	Yes
WHR	Waist/Hips	No
ABSI	$Waist \times Height^{5/6} / Weight^{2/3}$	No
Urea	Urea Level in Serum	Yes
Creatinine	Creatinine Level in Serum	Yes
Glucose	Fasing Glucose Level in Blood	Yes
TC	Total Cholesterol Level in Serum	Yes
HDL	HDL Cholesterol Level in serum	Yes
SBP	Systolic Blood Pressure	Yes
DBP	Diastolic Blood Pressure	Yes
HR	Heart Rate	Yes
Age	Volunteer Age at Clinic Appointment	Yes
Sex	Sex of Volunteer	Yes
SIMD	Scottish Index of Multiple Deprivation[e]	Yes
Centre	9 Clinics where phenotypes are measured	Yes
20PCs	The first 20 Principal Components of GRM_g	No

[e.] SIMD is a ranking based on environment, income, education etc. for each 350 household area [38].

Name	Unit	Natural logarithm transformation	Mean(s.d.)[f]
Height	cm	No	167.62(9.52)
Weight	kg	Yes	4.31(0.21)
Fat	percentage	No	30.73(9.35)
BMI	kg/cm^2	Yes	3.28(0.18)
Hips	cm	Yes	4.64(0.10)
Waist	cm	Yes	4.49(0.15)
WHR	(None)	Yes	-0.15(0.10)
ABSI	$cm^{11/6}$ / $kg^{2/3}$	Yes	-2.56(0.07)
Urea	mmol/l	Yes	1.63(0.27)
Creatinine	µmol/l	Yes	4.28(0.19)
Glucose	mmol/l	Yes	1.56(0.12)
TC	mmol/l	Yes	1.63(0.21)
HDL	mmol/l	Yes	0.35(0.28)
SBP	mmHg	Yes	4.87(0.13)
DBP	mmHg	Yes	4.38(0.13)
HR	beats/min	Yes	4.23(0.16)
Age	years old	No	52.20(13.64)
Sex	(F - female, M - male)	No	F: 5788; M:4075
SIMD	(None)	No	3994.13(1830.39)
Centre	(None)	No	(None)
20PCs	(None)	No	(None)

[f]·The mean and s.d. are calculated after transformation (where applicable) and for sex, the values are the number of individuals for each sex.

Name	Sex		Age		Sex-Age	
	Var%[g]	Effect[h]	Var%	Effect	Var%	Effect
Height	51.13%	-1.34E+01	4.43%	-1.43E-01	0.00%[NS]	-7.54E-03[NS]
Weight	19.83%	-1.70E-01	0.03%	9.16E-04[NS]	0.01%[NS]	-3.60E-04[NS]
Fat	38.67%	1.42E+01	5.35%	2.40E-01	0.11%	-4.66E-02
BMI	0.51%	-1.92E-02[NS]	2.45%	2.41E-03	0.00%[NS]	-1.36E-04[NS]
Hips	0.00%[NS]	5.37E-03[NS]	0.37%	6.90E-04	0.01%[NS]	-1.26E-04[NS]
Waist	13.96%	-9.34E-02	4.15%	3.11E-03	0.03%[NS]	-4.21E-04[NS]
WHR	30.40%	-9.76E-02	6.06%	2.43E-03	0.04%	-3.10E-04
ABSI	14.69%	-4.27E-02	5.85%	1.86E-03	0.07%	-2.86E-04
Urea	4.00%	-2.93E-01	14.71%	1.80E-03	0.83%	3.64E-03
Creatinine	31.08%	-2.07E-01	1.50%	2.04E-03	0.00%[NS]	-1.51E-04[NS]
Glucose	2.21%	-6.05E-02	7.39%	1.75E-03	0.07%	4.83E-04
TC	1.28%	-1.61E-01	3.58%	-3.47E-03	1.65%	4.05E-03
HDL	14.28%	1.28E-01	0.70%	-1.28E-03	0.17%	1.72E-03
SBP	4.19%	-1.72E-01	16.19%	5.32E-04[NS]	1.24%	2.26E-03
DBP	2.53%	-4.89E-02	3.49%	1.54E-03	0.01%[NS]	1.71E-04[NS]
HR	2.08%	7.11E-02	0.37%	1.05E-04[NS]	0.04%[NS]	-4.70E-04[NS]
Age	(None)	(None)	(None)	(None)	(None)	(None)
Sex	(None)	(None)	(None)	(None)	(None)	(None)
SIMD	(None)	(None)	(None)	(None)	(None)	(None)
Centre	(None)	(None)	(None)	(None)	(None)	(None)
20PCs	(None)	(None)	(None)	(None)	(None)	(None)

[g.] Var%: the proportion of phenotypic variance explained by each covariate neglecting genetic factors, estimated using ANOVA: Phenotype ~ Sex + Age + Sex-Age + SIMD + Clinics + 20PCs.

[h.] Effect: the estimated effect of each covariate neglecting genetic factors, estimated using ANOVA: Phenotype ~ Sex + Age + Sex-Age + SIMD + Centre + 20PCs.

[NS.] Not significant. The estimate is non-significant due to p-value > 0.05. F-test for variances and t-test for effects.

Name	SIMD		Centre	20PCs
	Var%	Effect	Var%	Var%
Height	0.83%	4.51E-04	0.14%	0.72%
Weight	0.35%	-7.01E-06	0.28%	0.54%
Fat	0.79%	-4.39E-04	0.46%	0.37%
BMI	1.62%	-1.23E-05	0.38%	0.52%
Hips	0.60%	-4.34E-06	1.03%	0.57%
Waist	1.48%	-1.02E-05	0.31%	0.44%
WHR	1.08%	-5.53E-06	0.25%	0.26%
ABSI	0.71%	-3.43E-06	1.84%	0.43%
Urea	0.05%	3.26E-06	0.27%	0.22% NS
Creatinine	0.00% NS	-7.03E-07 NS	0.68%	0.25%
Glucose	0.14%	-2.23E-06	0.53%	0.34%
TC	0.36%	6.55E-06	0.29%	0.34%
HDL	1.00%	1.57E-05	1.98%	0.56%
SBP	0.10%	-2.55E-06	0.41%	0.18% NS
DBP	0.01% NS	-1.15E-06 NS	1.05%	0.48%
HR	0.50%	-6.90E-06	0.67%	0.42%
Age	(None)	(None)	(None)	(None)
Sex	(None)	(None)	(None)	(None)
SIMD	(None)	(None)	(None)	(None)
Centre	(None)	(None)	(None)	(None)
20PCs	(None)	(None)	(None)	(None)

Table S2.2

Table S2.2 Abbreviations and equations for terms for all 31 possible alternative models used in our study

Model	GRM_g	GRM_{kin}	ERM_{Family}	ERM_{Sib}	ERM_{Couple}
G	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\varepsilon^2}$				
K		$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_{kin}^2 + \sigma_\varepsilon^2}$			
F			$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_{ef}^2 + \sigma_\varepsilon^2}$		
S				$e_s^2 = \frac{\sigma_{es}^2}{\sigma_{es}^2 + \sigma_\varepsilon^2}$	
C					$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_{ec}^2 + \sigma_\varepsilon^2}$
GK	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_\varepsilon^2}$	$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_\varepsilon^2}$			
GF	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_\varepsilon^2}$		$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_\varepsilon^2}$		

Model	GRM_g	GRM_{kin}	ERM_{Family}	ERM_{Sib}	ERM_{Couple}
GS	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$			$e_s^2 = \frac{\sigma_{es}^2}{\sigma_g^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	
GC	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$				$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_g^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
KF		$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_\varepsilon^2}$	$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_\varepsilon^2}$		
KS		$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_{kin}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$		$e_s^2 = \frac{\sigma_{es}^2}{\sigma_{kin}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	
KC		$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_{kin}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$			$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_{kin}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
FS			$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	$e_s^2 = \frac{\sigma_{es}^2}{\sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	
FC			$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$		$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
SC				$e_s^2 = \frac{\sigma_{es}^2}{\sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$

Model	GRM_g	GRM_{kin}	ERM_{Family}	ERM_{Sib}	ERM_{Couple}
GKF	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_\varepsilon^2}$	$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_\varepsilon^2}$	$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_\varepsilon^2}$		
GKS	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$		$e_S^2 = \frac{\sigma_{es}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	
GKC	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$			$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_g^2 + \sigma_{gkin}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
GFS	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$		$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	$e_s^2 = \frac{\sigma_{es}^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	
GFC	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$		$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$		$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
GSC	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$			$e_s^2 = \frac{\sigma_{es}^2}{\sigma_g^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_g^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
KFS		$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	$e_s^2 = \frac{\sigma_{es}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	
KFC		$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$		$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$

Model	GRM_g	GRM_{kin}	ERM_{Family}	ERM_{Sib}	ERM_{Couple}
KSC		$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_{kin}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$		$e_s^2 = \frac{\sigma_{es}^2}{\sigma_{kin}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_{kin}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
FSC			$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_s^2 = \frac{\sigma_{es}^2}{\sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
GKFS	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	$e_s^2 = \frac{\sigma_{es}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_\varepsilon^2}$	
GKFC	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$		$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
GKSC	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$		$e_s^2 = \frac{\sigma_{es}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
GFSC	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$		$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_s^2 = \frac{\sigma_{es}^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_g^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
KFSC		$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_s^2 = \frac{\sigma_{es}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$	$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2 + \sigma_\varepsilon^2}$
GKFSC	$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2}$	$h_{kin}^2 = \frac{\sigma_{kin}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2}$	$e_f^2 = \frac{\sigma_{ef}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2}$	$e_s^2 = \frac{\sigma_{es}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2}$	$e_c^2 = \frac{\sigma_{ec}^2}{\sigma_g^2 + \sigma_{kin}^2 + \sigma_{ef}^2 + \sigma_{es}^2 + \sigma_{ec}^2}$

Table S2.3

Table S2.3 Heritability estimates using model 'G' in subpopulations of GS10K with different GRM cut-offs

i^{th} degree of relatives	⩾6	≥5	≥4	≥3	≥2	≥1
GRM cut-off points	0.022	0.044	0.088	0.176	0.334	0.622
Trait	h^2_g **(s.e.)**	h^2_g **(s.e.)**	h^2_g **(s.e.)**	h^2_g **(s.e.)**	h^2_g **(s.e.)**	h^2_g **(s.e.)**
Height	0.54(0.06)	0.58(0.05)	0.57(0.05)	0.58(0.05)	0.61(0.04)	0.76(0.02)
Weight	0.27(0.06)	0.32(0.05)	0.3(0.05)	0.30(0.05)	0.31(0.04)	0.45(0.03)
Fat	0.28(0.06)	0.28(0.05)	0.27(0.05)	0.27(0.05)	0.29(0.05)	0.38(0.03)
BMI	0.24(0.06)	0.27(0.05)	0.26(0.05)	0.26(0.05)	0.27(0.04)	0.41(0.03)
Hips	0.23(0.06)	0.27(0.05)	0.23(0.05)	0.24(0.05)	0.24(0.05)	0.34(0.03)
Waist	0.17(0.06)	0.23(0.05)	0.21(0.05)	0.20(0.05)	0.20(0.05)	0.33(0.03)
WHR	0.13(0.06)	0.21(0.05)	0.20(0.05)	0.18(0.05)	0.17(0.05)	0.25(0.03)
ABSI	0.12(0.06)	0.14(0.05)	0.12(0.05)	0.09(0.05)	0.12(0.04)	0.19(0.03)
Urea	0.12(0.05)	0.09(0.05)	0.11(0.05)	0.12(0.05)	0.13(0.04)	0.21(0.03)
Creatinine	0.25(0.06)	0.23(0.05)	0.23(0.05)	0.22(0.05)	0.24(0.04)	0.45(0.03)
Glucose	0.21(0.06)	0.17(0.05)	0.18(0.05)	0.18(0.05)	0.20(0.05)	0.19(0.03)
TC	0.14(0.06)	0.15(0.05)	0.12(0.05)	0.13(0.05)	0.13(0.04)	0.27(0.03)
HDL	0.25(0.06)	0.30(0.05)	0.29(0.05)	0.30(0.05)	0.31(0.04)	0.42(0.03)
SBP	0.16(0.06)	0.17(0.05)	0.18(0.05)	0.18(0.05)	0.18(0.04)	0.20(0.03)
DBP	0.13(0.06)	0.15(0.05)	0.16(0.05)	0.15(0.05)	0.17(0.04)	0.17(0.03)
HR	0.09(0.06)	0.12(0.05)	0.13(0.05)	0.15(0.05)	0.11(0.04)	0.23(0.03)

[NS] Not significant. That variance component is non-significant according to LRT with p-value > 0.05.

Table S2.4

Table S2.4 (1) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'G'

Trait	Model: G[a]							
	σ^2_g (s.e.)[b]		σ^2_ε (s.e.)[c]		V (s.e.)[d]		h^2_g (s.e.)[e]	
Height	29.5133	(1.0786)	9.3316	(0.7767)	38.8449	(0.6166)	0.76	(0.02)
Weight	0.0153	(0.0010)	0.0187	(0.0009)	0.0340	(0.0005)	0.45	(0.03)
Fat	17.9608	(1.4184)	28.9508	(1.3089)	46.9116	(0.7211)	0.38	(0.03)
BMI	0.0121	(0.0009)	0.0175	(0.0008)	0.0296	(0.0005)	0.41	(0.03)
Hips	0.0029	(0.0003)	0.0058	(0.0002)	0.0087	(0.0001)	0.34	(0.03)
Waist	0.0060	(0.0005)	0.0123	(0.0005)	0.0183	(0.0003)	0.33	(0.03)
WHR	0.0016	(0.0002)	0.0048	(0.0002)	0.0064	(0.0001)	0.25	(0.03)
ABSI	0.0008	(0.0001)	0.0034	(0.0001)	0.0042	(0.0001)	0.19	(0.03)
Urea	0.0119	(0.0016)	0.0446	(0.0016)	0.0565	(0.0008)	0.21	(0.03)
Creatinine	0.0110	(0.0007)	0.0132	(0.0006)	0.0241	(0.0004)	0.45	(0.03)
Glucose	0.0024	(0.0004)	0.0107	(0.0004)	0.0131	(0.0002)	0.19	(0.03)
TC	0.0102	(0.0011)	0.0279	(0.0011)	0.0381	(0.0006)	0.27	(0.03)
HDL	0.0267	(0.0019)	0.0368	(0.0017)	0.0636	(0.0010)	0.42	(0.03)
SBP	0.0028	(0.0004)	0.0114	(0.0004)	0.0142	(0.0002)	0.20	(0.03)
DBP	0.0024	(0.0004)	0.0120	(0.0004)	0.0145	(0.0002)	0.17	(0.03)
HR	0.0060	(0.0008)	0.0199	(0.0008)	0.0259	(0.0004)	0.23	(0.03)

[a] Model '**G**' = $\mathbf{GRM}_g$

[b] This column shows the variance captured by matrix $\mathbf{GRM}_g$

[c] This column shows the residual variance

[d] This column shows the total phenotypic variance

[e] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM}_g$

Trait	Model: G				
		h^2_{gkin} (s.e.)[f]	$\%V_C$[g]	logL[h]	n[i]
Height	0.76	(0.02)	75.98%	-20858.40	9,150
Weight	0.45	(0.03)	45.00%	10903.20	9,118
Fat	0.38	(0.03)	38.29%	-21503.05	8,926
BMI	0.41	(0.03)	40.88%	11500.83	9,107
Hips	0.34	(0.03)	33.33%	16782.72	8,984
Waist	0.33	(0.03)	32.79%	13499.82	9,016
WHR	0.25	(0.03)	25.00%	18147.64	8,995
ABSI	0.19	(0.03)	19.05%	19935.63	8,962
Urea	0.21	(0.03)	21.06%	8517.75	9,148
Creatinine	0.45	(0.03)	45.64%	12503.98	9,146
Glucose	0.19	(0.03)	18.32%	14819.44	8,936
TC	0.27	(0.03)	26.77%	10316.20	9,136
HDL	0.42	(0.03)	41.98%	8053.25	9,125
SBP	0.20	(0.03)	19.72%	14805.64	9,144
DBP	0.17	(0.03)	16.55%	14700.05	9,141
HR	0.23	(0.03)	23.17%	12053.87	9,126

[f] This column shows the total heritability estimate, which is the sum of h^2_g and h^2_{kin}

[g] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[h] This column shows the log likelihood ratio for each analysis

[i] This column shows the number of records used for each analysis

Table S2.4 (2) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'K'

Trait	Model: K[a]							
	σ^2_{kin}(s.e.)[b]		σ^2_{ε} (s.e.)[c]		V (s.e.)[d]		h^2_{kin}(s.e.)[e]	
Height	32.4375	(1.1120)	5.9904	(0.8485)	38.4279	(0.5910)	0.84	(0.02)
Weight	0.0190	(0.0012)	0.0148	(0.0011)	0.0339	(0.0005)	0.56	(0.03)
Fat	21.6101	(1.7359)	25.1407	(1.6244)	46.7507	(0.7123)	0.46	(0.04)
BMI	0.0153	(0.0011)	0.0142	(0.0010)	0.0295	(0.0004)	0.52	(0.03)
Hips	0.0038	(0.0003)	0.0049	(0.0003)	0.0087	(0.0001)	0.44	(0.04)
Waist	0.0084	(0.0007)	0.0099	(0.0006)	0.0183	(0.0003)	0.46	(0.04)
WHR	0.0020	(0.0002)	0.0043	(0.0002)	0.0064	(0.0001)	0.32	(0.04)
ABSI	0.0012	(0.0002)	0.0030	(0.0002)	0.0042	(0.0001)	0.27	(0.04)
Urea	0.0147	(0.0021)	0.0418	(0.0021)	0.0564	(0.0008)	0.26	(0.04)
Creatinine	0.0146	(0.0008)	0.0094	(0.0007)	0.0240	(0.0004)	0.61	(0.03)
Glucose	0.0022	(0.0005)	0.0109	(0.0005)	0.0131	(0.0002)	0.17	(0.04)
TC	0.0138	(0.0014)	0.0242	(0.0014)	0.0380	(0.0006)	0.36	(0.04)
HDL	0.0319	(0.0022)	0.0314	(0.0021)	0.0634	(0.0010)	0.50	(0.03)
SBP	0.0034	(0.0005)	0.0107	(0.0005)	0.0142	(0.0002)	0.24	(0.04)
DBP	0.0028	(0.0005)	0.0117	(0.0006)	0.0145	(0.0002)	0.19	(0.04)
HR	0.0078	(0.0010)	0.0180	(0.0010)	0.0258	(0.0004)	0.30	(0.04)

[a] Model '**K**' = $\mathbf{GRM}_{kin}$

[b] This column shows the variance captured by matrix $\mathbf{GRM}_{kin}$

[c] This column shows the residual variance

[d] This column shows the total phenotypic variance

[e] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM}_{kin}$

Trait	Model: K				
	h^2_{gkin} (s.e.)[f]		%V_C[g]	logL[h]	n[i]
Height	0.84	(0.02)	84.41%	-20948.87	9,150
Weight	0.56	(0.03)	56.05%	10894.06	9,118
Fat	0.46	(0.04)	46.22%	-21516.87	8,926
BMI	0.52	(0.03)	51.86%	11493.50	9,107
Hips	0.44	(0.04)	43.68%	16779.11	8,984
Waist	0.46	(0.04)	45.90%	13507.43	9,016
WHR	0.32	(0.04)	31.25%	18145.39	8,995
ABSI	0.27	(0.04)	28.57%	19938.29	8,962
Urea	0.26	(0.04)	26.06%	8511.19	9,148
Creatinine	0.61	(0.03)	60.83%	12513.01	9,146
Glucose	0.17	(0.04)	16.79%	14806.81	8,936
TC	0.36	(0.04)	36.32%	10317.70	9,136
HDL	0.50	(0.03)	50.32%	8033.09	9,125
SBP	0.24	(0.04)	23.94%	14800.18	9,144
DBP	0.19	(0.04)	19.31%	14694.13	9,141
HR	0.30	(0.04)	30.23%	12051.04	9,126

[f] This column shows the total heritability estimate, which is the sum of h^2_g and h^2_{kin}

[g] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[h] This column shows the log likelihood ratio for each analysis

[i] This column shows the number of records used for each analysis

Table S2.4 (3) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'F'

Trait	Model: F[a]							
	σ^2_{ef} (s.e.)[b]		σ^2_{ε} (s.e.)[c]		V (s.e.)[d]		e^2_f (s.e.)[e]	
Height	16.1151	(0.6841)	22.6906	(0.5590)	38.8056	(0.6068)	0.42	(0.01)
Weight	0.0094	(0.0006)	0.0246	(0.0006)	0.0340	(0.0005)	0.28	(0.02)
Fat	11.1861	(0.9047)	35.7142	(0.9072)	46.9003	(0.7207)	0.24	(0.02)
BMI	0.0080	(0.0006)	0.0216	(0.0005)	0.0296	(0.0005)	0.27	(0.02)
Hips	0.0019	(0.0002)	0.0068	(0.0002)	0.0087	(0.0001)	0.22	(0.02)
Waist	0.0044	(0.0004)	0.0140	(0.0004)	0.0183	(0.0003)	0.24	(0.02)
WHR	0.0009	(0.0001)	0.0055	(0.0001)	0.0064	(0.0001)	0.15	(0.02)
ABSI	0.0005	(0.0001)	0.0037	(0.0001)	0.0042	(0.0001)	0.12	(0.02)
Urea	0.0081	(0.0010)	0.0484	(0.0012)	0.0565	(0.0008)	0.14	(0.02)
Creatinine	0.0067	(0.0005)	0.0174	(0.0004)	0.0241	(0.0004)	0.28	(0.02)
Glucose	0.0010	(0.0002)	0.0121	(0.0003)	0.0131	(0.0002)	0.07	(0.02)
TC	0.0062	(0.0007)	0.0319	(0.0008)	0.0381	(0.0006)	0.16	(0.02)
HDL	0.0147	(0.0012)	0.0488	(0.0012)	0.0635	(0.0010)	0.23	(0.02)
SBP	0.0017	(0.0003)	0.0125	(0.0003)	0.0142	(0.0002)	0.12	(0.02)
DBP	0.0015	(0.0003)	0.0130	(0.0003)	0.0145	(0.0002)	0.10	(0.02)
HR	0.0038	(0.0005)	0.0221	(0.0005)	0.0258	(0.0004)	0.15	(0.02)

[a] Model '**F**' = $\mathbf{ERM}_{\mathbf{Family}}$

[b] This column shows the variance captured by matrix $\mathbf{ERM}_{\mathbf{Family}}$

[c] This column shows the residual variance

[d] This column shows the total phenotypic variance

[e] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{\mathbf{Family}}$

Trait	Model: F		
	%V_C[f]	logL[g]	n[h]
Height	41.53%	-20971.18	9,150
Weight	27.73%	10899.26	9,118
Fat	23.85%	-21506.91	8,926
BMI	26.97%	11509.14	9,107
Hips	22.01%	16786.97	8,984
Waist	23.78%	13519.62	9,016
WHR	14.68%	18143.34	8,995
ABSI	11.55%	19934.01	8,962
Urea	14.29%	8520.39	9,148
Creatinine	27.61%	12501.02	9,146
Glucose	7.48%	14807.42	8,936
TC	16.23%	10313.36	9,136
HDL	23.15%	8027.90	9,125
SBP	12.06%	14806.06	9,144
DBP	10.14%	14699.05	9,141
HR	14.60%	12055.96	9,126

[f] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[g] This column shows the log likelihood ratio for each analysis

[h] This column shows the number of records used for each analysis

Table S2.4 (4) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'S'

Trait	Model: S[a]							
	σ^2_{es} (s.e.)[b]		σ^2_{ε} (s.e.)[c]		V (s.e.)[d]		e^2_s (s.e.)[e]	
Height	19.5744	(1.2700)	19.2443	(1.1807)	38.8187	(0.5821)	0.50	(0.03)
Weight	0.0101	(0.0014)	0.0238	(0.0014)	0.0339	(0.0005)	0.30	(0.04)
Fat	12.5497	(1.9687)	34.2239	(1.9624)	46.7736	(0.7046)	0.27	(0.04)
BMI	0.0080	(0.0012)	0.0215	(0.0012)	0.0295	(0.0004)	0.27	(0.04)
Hips	0.0018	(0.0004)	0.0068	(0.0004)	0.0087	(0.0001)	0.21	(0.04)
Waist	0.0043	(0.0008)	0.0140	(0.0008)	0.0182	(0.0003)	0.23	(0.04)
WHR	0.0009	(0.0003)	0.0054	(0.0003)	0.0064	(0.0001)	0.15	(0.04)
ABSI	0.0005	(0.0002)	0.0037	(0.0002)	0.0042	(0.0001)	0.11	(0.04)
Urea	0.0084	(0.0026)	0.0480	(0.0026)	0.0564	(0.0008)	0.15	(0.05)
Creatinine	0.0099	(0.0009)	0.0142	(0.0009)	0.0241	(0.0004)	0.41	(0.04)
Glucose	0.0022	(0.0007)	0.0109	(0.0007)	0.0131	(0.0002)	0.17	(0.05)
TC	0.0109	(0.0015)	0.0271	(0.0015)	0.0381	(0.0006)	0.29	(0.04)
HDL	0.0189	(0.0025)	0.0445	(0.0025)	0.0635	(0.0009)	0.30	(0.04)
SBP	0.0029	(0.0007)	0.0113	(0.0007)	0.0142	(0.0002)	0.20	(0.05)
DBP	0.0019	(0.0006)	0.0125	(0.0007)	0.0145	(0.0002)	0.13	(0.04)
HR	0.0044	(0.0012)	0.0214	(0.0012)	0.0258	(0.0004)	0.17	(0.05)

[a] Model '**S**' = $\mathbf{ERM}_{Sib}$

[b] This column shows the variance captured by matrix $\mathbf{ERM}_{Sib}$

[c] This column shows the residual variance

[d] This column shows the total phenotypic variance

[e] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{Sib}$

Trait	Model: S		
	%V_C[f]	logL[g]	n[h]
Height	50.43%	-21212.44	9,150
Weight	29.66%	10797.67	9,118
Fat	26.83%	-21576.62	8,926
BMI	27.02%	11412.19	9,107
Hips	21.09%	16725.18	8,984
Waist	23.47%	13446.72	9,016
WHR	14.59%	18114.48	8,995
ABSI	11.28%	19915.42	8,962
Urea	14.91%	8491.80	9,148
Creatinine	41.01%	12417.21	9,146
Glucose	16.68%	14801.87	8,936
TC	28.72%	10290.17	9,136
HDL	29.85%	7954.10	9,125
SBP	20.31%	14787.38	9,144
DBP	13.34%	14684.55	9,141
HR	17.10%	12024.89	9,126

[f] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[g] This column shows the log likelihood ratio for each analysis

[h] This column shows the number of records used for each analysis

Table S2.4 (5) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'C'

Trait	Model: C[a]							
	σ^2_{ec} (s.e.)[b]		σ^2_{ε} (s.e.)[c]		V (s.e.)[d]		e^2_c (s.e.)[e]	
Height	10.6492	(1.0885)	28.1385	(1.0808)	38.7878	(0.5801)	0.27	(0.03)
Weight	0.0066	(0.0010)	0.0273	(0.0010)	0.0339	(0.0005)	0.19	(0.03)
Fat	8.9600	(1.4296)	37.8227	(1.4677)	46.7827	(0.7051)	0.19	(0.03)
BMI	0.0066	(0.0009)	0.0229	(0.0009)	0.0295	(0.0004)	0.22	(0.03)
Hips	0.0015	(0.0003)	0.0072	(0.0003)	0.0087	(0.0001)	0.17	(0.03)
Waist	0.0039	(0.0005)	0.0144	(0.0006)	0.0183	(0.0003)	0.21	(0.03)
WHR	0.0006	(0.0002)	0.0058	(0.0002)	0.0064	(0.0001)	0.09	(0.03)
ABSI	0.0002	$(0.0001)^{NS}$	0.0040	(0.0001)	0.0042	(0.0001)	0.05	$(0.03)^{NS}$
Urea	0.0072	(0.0017)	0.0492	(0.0018)	0.0564	(0.0008)	0.13	(0.03)
Creatinine	0.0032	(0.0007)	0.0207	(0.0007)	0.0240	(0.0004)	0.13	(0.03)
Glucose	0.0007	(0.0004)	0.0124	(0.0004)	0.0131	(0.0002)	0.05	(0.03)
TC	0.0026	(0.0011)	0.0355	(0.0012)	0.0380	(0.0006)	0.07	(0.03)
HDL	0.0097	(0.0018)	0.0538	(0.0019)	0.0635	(0.0009)	0.15	(0.03)
SBP	0.0014	(0.0004)	0.0127	(0.0004)	0.0141	(0.0002)	0.10	(0.03)
DBP	0.0014	(0.0004)	0.0131	(0.0005)	0.0145	(0.0002)	0.10	(0.03)
HR	0.0024	(0.0008)	0.0234	(0.0008)	0.0258	(0.0004)	0.09	(0.03)

[a] Model '**C**' = $\mathbf{ERM}_{\mathbf{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{ERM}_{\mathbf{Couple}}$

[c] This column shows the residual variance

[d] This column shows the total phenotypic variance

[e] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{\mathbf{Couple}}$

Trait	Model: C		
	%V_C[f]	logL[g]	n[h]
Height	27.46%	-21241.61	9,150
Weight	19.47%	10794.77	9,118
Fat	19.15%	-21577.47	8,926
BMI	22.37%	11418.82	9,107
Hips	17.24%	16727.49	8,984
Waist	21.31%	13454.62	9,016
WHR	9.38%	18112.60	8,995
ABSI	4.76%	19913.65	8,962
Urea	12.77%	8495.77	9,148
Creatinine	13.33%	12393.33	9,146
Glucose	5.34%	14799.98	8,936
TC	6.84%	10270.03	9,136
HDL	15.28%	7942.22	9,125
SBP	9.93%	14787.34	9,144
DBP	9.66%	14685.18	9,141
HR	9.30%	12024.34	9,126

[f] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[g] This column shows the log likelihood ratio for each analysis

[h] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (6) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GK'

Trait	Model: GK[a]							
	σ^2_g (s.e.)[b]		σ^2_{kin}(s.e.)[c]		σ^2_ε (s.e.)[d]		V (s.e.)[e]	
Height	21.4798	(1.5917)	10.8291	(1.7221)	6.1236	(0.8222)	38.4325	(0.6033)
Weight	0.0095	(0.0014)	0.0097	(0.0017)	0.0148	(0.0011)	0.0339	(0.0005)
Fat	12.4104	(1.9812)	9.3921	(2.5017)	25.0282	(1.6086)	46.8306	(0.7184)
BMI	0.0075	(0.0012)	0.0079	(0.0015)	0.0142	(0.0010)	0.0295	(0.0005)
Hips	0.0018	(0.0004)	0.0020	(0.0005)	0.0049	(0.0003)	0.0087	(0.0001)
Waist	0.0029	(0.0007)	0.0055	(0.0010)	0.0099	(0.0006)	0.0183	(0.0003)
WHR	0.0010	(0.0003)	0.0011	(0.0003)	0.0043	(0.0002)	0.0064	(0.0001)
ABSI	0.0004	(0.0002)	0.0008	(0.0002)	0.0030	(0.0002)	0.0042	(0.0001)
Urea	0.0088	(0.0022)	0.0058	(0.0030)	0.0418	(0.0021)	0.0565	(0.0008)
Creatinine	0.0057	(0.0010)	0.0090	(0.0012)	0.0094	(0.0007)	0.0241	(0.0004)
Glucose	0.0024	(0.0005)	0.0000	(0.0007) [NS]	0.0107	(0.0005)	0.0131	(0.0002)
TC	0.0056	(0.0015)	0.0083	(0.0020)	0.0241	(0.0014)	0.0381	(0.0006)
HDL	0.0188	(0.0026)	0.0132	(0.0033)	0.0315	(0.0021)	0.0634	(0.0010)
SBP	0.0021	(0.0006)	0.0014	(0.0008)	0.0107	(0.0005)	0.0142	(0.0002)
DBP	0.0020	(0.0006)	0.0008	(0.0008) [NS]	0.0117	(0.0006)	0.0145	(0.0002)
HR	0.0038	(0.0010)	0.0040	(0.0014)	0.0180	(0.0010)	0.0259	(0.0004)

[a] Model '**GK**' = $\mathbf{GRM_g} + \mathbf{GRM_{kin}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[d] This column shows the residual variance

[e] This column shows the total phenotypic variance

Trait	Model: GK								
	h^2_g (s.e.)[f]		h^2_{kin} (s.e.)[g]		h^2_{gkin} (s.e.)[h]		%V_C[i]	logL[j]	n[k]
Height	0.56	(0.04)	0.28	(0.04)	0.84	(0.04)	84.07%	-20837.78	9,150
Weight	0.28	(0.04)	0.29	(0.05)	0.57	(0.05)	56.54%	10919.02	9,118
Fat	0.27	(0.04)	0.20	(0.05)	0.47	(0.05)	46.56%	-21495.95	8,926
BMI	0.25	(0.04)	0.27	(0.05)	0.52	(0.05)	51.95%	11514.21	9,107
Hips	0.21	(0.04)	0.23	(0.05)	0.44	(0.05)	44.01%	16792.24	8,984
Waist	0.16	(0.04)	0.30	(0.05)	0.46	(0.05)	45.89%	13515.28	9,016
WHR	0.15	(0.04)	0.17	(0.05)	0.32	(0.05)	32.32%	18152.66	8,995
ABSI	0.10	(0.04)	0.18	(0.05)	0.28	(0.05)	27.70%	19941.24	8,962
Urea	0.16	(0.04)	0.10	(0.05)	0.26	(0.05)	25.91%	8519.62	9,148
Creatinine	0.24	(0.04)	0.37	(0.05)	0.61	(0.05)	60.98%	12531.16	9,146
Glucose	0.19	(0.04)	0.00	(0.06) [NS]	0.19	(0.05)	18.53%	14819.44	8,936
TC	0.15	(0.04)	0.22	(0.05)	0.37	(0.05)	36.64%	10325.29	9,136
HDL	0.30	(0.04)	0.21	(0.05)	0.50	(0.05)	50.42%	8061.69	9,125
SBP	0.15	(0.04)	0.10	(0.05)	0.25	(0.05)	24.63%	14807.35	9,144
DBP	0.14	(0.04)	0.05	(0.05) [NS]	0.19	(0.05)	19.47%	14700.60	9,141
HR	0.15	(0.04)	0.16	(0.05)	0.30	(0.05)	30.38%	12058.15	9,126

[f] This column shows the proportion of total phenotypic variance captured by matrix **GRM_g**

[g] This column shows the proportion of total phenotypic variance captured by matrix **GRM_{kin}**

[h] This column shows the total heritability estimate, which is the sum of h^2_g and h^2_{kin}

[i] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[j] This column shows the log likelihood ratio for each analysis

[k] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (7) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GF'

Trait	Model: GF[a]							
	σ^2_g (s.e.)[b]		σ^2_{ef} (s.e.)[c]		σ^2_ε (s.e.)[d]		V (s.e.)[e]	
Height	20.7700	(1.2374)	8.2316	(0.7922)	9.6687	(0.7264)	38.6703	(0.6119)
Weight	0.0096	(0.0012)	0.0061	(0.0007)	0.0184	(0.0009)	0.0341	(0.0005)
Fat	11.3064	(1.6392)	7.0575	(1.0616)	28.5714	(1.2780)	46.9353	(0.7247)
BMI	0.0070	(0.0010)	0.0055	(0.0007)	0.0171	(0.0008)	0.0296	(0.0005)
Hips	0.0018	(0.0003)	0.0013	(0.0002)	0.0056	(0.0002)	0.0087	(0.0001)
Waist	0.0030	(0.0006)	0.0033	(0.0004)	0.0120	(0.0005)	0.0183	(0.0003)
WHR	0.0011	(0.0002)	0.0006	(0.0001)	0.0048	(0.0002)	0.0064	(0.0001)
ABSI	0.0005	(0.0001)	0.0003	(0.0001)	0.0034	(0.0001)	0.0042	(0.0001)
Urea	0.0072	(0.0019)	0.0056	(0.0012)	0.0438	(0.0016)	0.0565	(0.0008)
Creatinine	0.0073	(0.0008)	0.0047	(0.0005)	0.0123	(0.0006)	0.0242	(0.0004)
Glucose	0.0022	(0.0005)	0.0003	(0.0003) [NS]	0.0106	(0.0004)	0.0131	(0.0002)
TC	0.0067	(0.0013)	0.0040	(0.0008)	0.0275	(0.0011)	0.0381	(0.0006)
HDL	0.0191	(0.0022)	0.0082	(0.0014)	0.0362	(0.0017)	0.0636	(0.0010)
SBP	0.0018	(0.0005)	0.0012	(0.0003)	0.0112	(0.0004)	0.0142	(0.0002)
DBP	0.0016	(0.0005)	0.0009	(0.0003)	0.0120	(0.0004)	0.0145	(0.0002)
HR	0.0036	(0.0009)	0.0026	(0.0006)	0.0196	(0.0008)	0.0259	(0.0004)

[a] Model '**GF**' = $\mathbf{GRM_g}$ + $\mathbf{ERM_{Family}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[d] This column shows the residual variance

[e] This column shows the total phenotypic variance

Trait	Model: GF								
	h^2_g (s.e.)[f]		e^2_f (s.e.)[g]		h^2_{gkin} (s.e.)[h]		%V_C[i]	logL[j]	n[k]
Height	0.54	(0.03)	0.21	(0.02)	0.54	(0.03)	75.00%	-20805.37	9,150
Weight	0.28	(0.03)	0.18	(0.02)	0.28	(0.03)	46.12%	10936.81	9,118
Fat	0.24	(0.03)	0.15	(0.02)	0.24	(0.03)	39.13%	-21481.57	8,926
BMI	0.24	(0.03)	0.19	(0.02)	0.24	(0.03)	42.30%	11536.02	9,107
Hips	0.20	(0.03)	0.15	(0.02)	0.20	(0.03)	35.32%	16805.38	8,984
Waist	0.16	(0.03)	0.18	(0.02)	0.16	(0.03)	34.49%	13531.90	9,016
WHR	0.17	(0.03)	0.09	(0.02)	0.17	(0.03)	25.17%	18154.97	8,995
ABSI	0.13	(0.03)	0.07	(0.02)	0.13	(0.03)	19.87%	19940.92	8,962
Urea	0.13	(0.03)	0.10	(0.02)	0.13	(0.03)	22.51%	8528.16	9,148
Creatinine	0.30	(0.03)	0.19	(0.02)	0.30	(0.03)	49.36%	12546.86	9,146
Glucose	0.17	(0.03)	0.02	(0.02) [NS]	0.17	(0.03)	18.94%	14820.07	8,936
TC	0.18	(0.03)	0.10	(0.02)	0.18	(0.03)	28.00%	10328.22	9,136
HDL	0.30	(0.03)	0.13	(0.02)	0.30	(0.03)	43.02%	8072.52	9,125
SBP	0.13	(0.03)	0.08	(0.02)	0.13	(0.03)	21.17%	14814.02	9,144
DBP	0.11	(0.03)	0.06	(0.02)	0.11	(0.03)	17.30%	14704.63	9,141
HR	0.14	(0.03)	0.10	(0.02)	0.14	(0.03)	24.06%	12064.95	9,126

[f] This column shows the proportion of total phenotypic variance captured by matrix **GRM_g**

[g] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Family}**

[h] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[i] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[j] This column shows the log likelihood ratio for each analysis

[k] This column shows the number of records used for each analysis

Table S2.4 (8) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GS'

Trait	Model: GS[a]							
	σ^2_g (s.e.)[b]		σ^2_{es} (s.e.)[c]		σ^2_ε (s.e.)[d]		V (s.e.)[e]	
Height	29.2833	(1.1042)	3.4149	(1.1535)	6.1969	(1.2558)	38.8951	(0.6193)
Weight	0.0150	(0.0010)	0.0026	(0.0013)	0.0165	(0.0014)	0.0340	(0.0005)
Fat	17.4182	(1.4604)	3.5725	(1.9486)	25.9267	(2.0605)	46.9174	(0.7215)
BMI	0.0118	(0.0009)	0.0021	(0.0012)	0.0157	(0.0013)	0.0296	(0.0005)
Hips	0.0029	(0.0003)	0.0004	(0.0004) [NS]	0.0054	(0.0004)	0.0087	(0.0001)
Waist	0.0058	(0.0006)	0.0013	(0.0008) [NS]	0.0112	(0.0008)	0.0183	(0.0003)
WHR	0.0015	(0.0002)	0.0001	(0.0003) [NS]	0.0047	(0.0003)	0.0064	(0.0001)
ABSI	0.0008	(0.0001)	0.0001	(0.0002) [NS]	0.0033	(0.0002)	0.0042	(0.0001)
Urea	0.0116	(0.0017)	0.0021	(0.0026) [NS]	0.0428	(0.0027)	0.0565	(0.0008)
Creatinine	0.0104	(0.0008)	0.0041	(0.0009)	0.0097	(0.0010)	0.0242	(0.0004)
Glucose	0.0024	(0.0004)	0.0011	(0.0007) [NS]	0.0096	(0.0007)	0.0131	(0.0002)
TC	0.0092	(0.0012)	0.0066	(0.0016)	0.0224	(0.0016)	0.0381	(0.0006)
HDL	0.0261	(0.0019)	0.0043	(0.0025)	0.0332	(0.0027)	0.0636	(0.0010)
SBP	0.0027	(0.0004)	0.0017	(0.0007)	0.0099	(0.0007)	0.0142	(0.0002)
DBP	0.0023	(0.0004)	0.0008	(0.0006) [NS]	0.0113	(0.0007)	0.0145	(0.0002)
HR	0.0058	(0.0008)	0.0013	(0.0012) [NS]	0.0187	(0.0013)	0.0259	(0.0004)

[a] Model '**GS**' = $\mathbf{GRM_g}$ + $\mathbf{ERM_{Sib}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

[d] This column shows the residual variance

[e] This column shows the total phenotypic variance

Trait	Model: GS								
	h^2_g (s.e.)[f]		e^2_s (s.e.)[g]		h^2_{gkin} (s.e.)[h]		%V_C[i]	logL[j]	n[k]
Height	0.75	(0.02)	0.09	(0.03)	0.75	(0.02)	84.07%	-20854.86	9,150
Weight	0.44	(0.03)	0.08	(0.04)	0.44	(0.03)	51.56%	10904.97	9,118
Fat	0.37	(0.03)	0.08	(0.04)	0.37	(0.03)	44.74%	-21501.50	8,926
BMI	0.40	(0.03)	0.07	(0.04)	0.40	(0.03)	46.92%	11502.29	9,107
Hips	0.33	(0.03)	0.05	(0.04) [NS]	0.33	(0.03)	37.47%	16783.49	8,984
Waist	0.32	(0.03)	0.07	(0.04) [NS]	0.32	(0.03)	38.80%	13501.04	9,016
WHR	0.24	(0.03)	0.02	(0.04) [NS]	0.24	(0.03)	25.77%	18147.72	8,995
ABSI	0.19	(0.03)	0.01	(0.05) [NS]	0.19	(0.03)	21.38%	19935.67	8,962
Urea	0.20	(0.03)	0.04	(0.05) [NS]	0.20	(0.03)	24.19%	8518.05	9,148
Creatinine	0.43	(0.03)	0.17	(0.04)	0.43	(0.03)	59.72%	12510.94	9,146
Glucose	0.18	(0.03)	0.08	(0.05) [NS]	0.18	(0.03)	26.37%	14820.42	8,936
TC	0.24	(0.03)	0.17	(0.04)	0.24	(0.03)	41.26%	10324.28	9,136
HDL	0.41	(0.03)	0.07	(0.04)	0.41	(0.03)	47.82%	8054.67	9,125
SBP	0.19	(0.03)	0.12	(0.05)	0.19	(0.03)	30.38%	14808.07	9,144
DBP	0.16	(0.03)	0.06	(0.04) [NS]	0.16	(0.03)	21.64%	14700.85	9,141
HR	0.22	(0.03)	0.05	(0.05) [NS]	0.22	(0.03)	27.45%	12054.35	9,126

[f] This column shows the proportion of total phenotypic variance captured by matrix **GRM_g**

[g] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Sib}**

[h] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[i] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[j] This column shows the log likelihood ratio for each analysis

[k] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (9) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GC'

Trait	Model: GC[a]							
	σ^2_g (s.e.)[b]		σ^2_{ec} (s.e.)[c]		σ^2_ε (s.e.)[d]		V (s.e.)[e]	
Height	30.9259	(1.1844)	6.1344	(0.9552)	1.8881	(1.4055)	38.9483	(0.6259)
Weight	0.0157	(0.0011)	0.0052	(0.0010)	0.0132	(0.0014)	0.0341	(0.0005)
Fat	18.5453	(1.4849)	7.8055	(1.3926)	20.6053	(1.9706)	46.9561	(0.7253)
BMI	0.0124	(0.0009)	0.0055	(0.0008)	0.0117	(0.0012)	0.0296	(0.0005)
Hips	0.0030	(0.0003)	0.0013	(0.0003)	0.0043	(0.0004)	0.0087	(0.0001)
Waist	0.0061	(0.0006)	0.0034	(0.0005)	0.0088	(0.0008)	0.0183	(0.0003)
WHR	0.0016	(0.0002)	0.0005	(0.0002)	0.0043	(0.0003)	0.0064	(0.0001)
ABSI	0.0008	(0.0001)	0.0002	$(0.0001)^{NS}$	0.0032	(0.0002)	0.0042	(0.0001)
Urea	0.0123	(0.0017)	0.0070	(0.0016)	0.0371	(0.0024)	0.0565	(0.0008)
Creatinine	0.0114	(0.0008)	0.0029	(0.0006)	0.0099	(0.0010)	0.0242	(0.0004)
Glucose	0.0025	(0.0004)	0.0007	(0.0004)	0.0099	(0.0006)	0.0131	(0.0002)
TC	0.0104	(0.0011)	0.0023	(0.0011)	0.0254	(0.0016)	0.0381	(0.0006)
HDL	0.0277	(0.0020)	0.0082	(0.0018)	0.0277	(0.0026)	0.0636	(0.0010)
SBP	0.0029	(0.0004)	0.0014	(0.0004)	0.0099	(0.0006)	0.0142	(0.0002)
DBP	0.0025	(0.0004)	0.0013	(0.0004)	0.0107	(0.0006)	0.0145	(0.0002)
HR	0.0061	(0.0008)	0.0022	(0.0007)	0.0176	(0.0011)	0.0259	(0.0004)

[a] Model '**GC**' = $\mathbf{GRM_g}$ + $\mathbf{ERM_{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Couple}}$

[d] This column shows the residual variance

[e] This column shows the total phenotypic variance

Trait	Model: GC								
	h^2_g (s.e.)[f]		e^2_c (s.e.)[g]		h^2_{gkin} (s.e.)[h]		%V_C[i]	logL[j]	n[k]
Height	0.79	(0.02)	0.16	(0.02)	0.79	(0.02)	95.15%	-20839.89	9,150
Weight	0.46	(0.03)	0.15	(0.03)	0.46	(0.03)	61.20%	10915.62	9,118
Fat	0.39	(0.03)	0.17	(0.03)	0.39	(0.03)	56.12%	-21489.81	8,926
BMI	0.42	(0.03)	0.19	(0.03)	0.42	(0.03)	60.46%	11519.00	9,107
Hips	0.35	(0.03)	0.15	(0.03)	0.35	(0.03)	50.11%	16794.44	8,984
Waist	0.33	(0.03)	0.19	(0.03)	0.33	(0.03)	52.04%	13516.74	9,016
WHR	0.25	(0.03)	0.08	(0.03)	0.25	(0.03)	32.58%	18150.47	8,995
ABSI	0.19	(0.03)	0.05	(0.03) [NS]	0.19	(0.03)	24.14%	19936.62	8,962
Urea	0.22	(0.03)	0.12	(0.03)	0.22	(0.03)	34.26%	8526.37	9,148
Creatinine	0.47	(0.03)	0.12	(0.03)	0.47	(0.03)	59.19%	12514.03	9,146
Glucose	0.19	(0.03)	0.05	(0.03)	0.19	(0.03)	24.25%	14821.30	8,936
TC	0.27	(0.03)	0.06	(0.03)	0.27	(0.03)	33.27%	10318.33	9,136
HDL	0.44	(0.03)	0.13	(0.03)	0.44	(0.03)	56.49%	8063.55	9,125
SBP	0.21	(0.03)	0.10	(0.03)	0.21	(0.03)	30.25%	14812.75	9,144
DBP	0.17	(0.03)	0.09	(0.03)	0.17	(0.03)	26.24%	14704.67	9,141
HR	0.23	(0.03)	0.08	(0.03)	0.23	(0.03)	31.82%	12057.99	9,126

[f] This column shows the proportion of total phenotypic variance captured by matrix **GRM_g**

[g] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Couple}**

[h] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[i] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[j] This column shows the log likelihood ratio for each analysis

[k] This column shows the number of records used for each analysis

Table S2.4 (10) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'KF'

Trait	Model: KF[a]							
	σ^2_{kin} (s.e.)[b]		σ^2_{ef} (s.e.)[c]		σ^2_{ε} (s.e.)[d]		V (s.e.)[e]	
Height	19.8456	(1.7450)	9.0618	(1.0002)	9.8502	(1.0247)	38.7576	(0.6073)
Weight	0.0098	(0.0019)	0.0060	(0.0010)	0.0182	(0.0013)	0.0340	(0.0005)
Fat	8.7921	(2.8241)	7.9779	(1.3815)	30.1126	(1.9485)	46.8825	(0.7204)
BMI	0.0060	(0.0017)	0.0059	(0.0008)	0.0177	(0.0012)	0.0296	(0.0005)
Hips	0.0016	(0.0005)	0.0013	(0.0003)	0.0057	(0.0004)	0.0087	(0.0001)
Waist	0.0032	(0.0011)	0.0032	(0.0005)	0.0119	(0.0008)	0.0183	(0.0003)
WHR	0.0013	(0.0004)	0.0005	(0.0002)	0.0046	(0.0003)	0.0064	(0.0001)
ABSI	0.0009	(0.0003)	0.0002	(0.0001)[NS]	0.0032	(0.0002)	0.0042	(0.0001)
Urea	0.0028	(0.0035)[NS]	0.0071	(0.0016)	0.0466	(0.0025)	0.0565	(0.0008)
Creatinine	0.0095	(0.0012)	0.0037	(0.0007)	0.0109	(0.0008)	0.0242	(0.0004)
Glucose	0.0012	(0.0008)[NS]	0.0006	(0.0004)	0.0113	(0.0006)	0.0131	(0.0002)
TC	0.0091	(0.0023)	0.0030	(0.0011)	0.0260	(0.0016)	0.0381	(0.0006)
HDL	0.0198	(0.0035)	0.0078	(0.0018)	0.0359	(0.0024)	0.0635	(0.0001)
SBP	0.0012	(0.0008)[NS]	0.0013	(0.0004)	0.0116	(0.0006)	0.0142	(0.0002)
DBP	0.0005	(0.0009)[NS]	0.0013	(0.0004)	0.0127	(0.0007)	0.0145	(0.0002)
HR	0.0032	(0.0016)	0.0027	(0.0007)	0.0199	(0.0011)	0.0259	(0.0004)

[a] Model '**KF**' = $\mathbf{GRM}_{kin}$ + $\mathbf{ERM}_{Family}$

[b] This column shows the variance captured by matrix $\mathbf{GRM}_{kin}$

[c] This column shows the variance captured by matrix $\mathbf{ERM}_{Family}$

[d] This column shows the residual variance

[e] This column shows the total phenotypic variance

Trait	Model: KF								
	h^2_{kin} (s.e.)[f]		e^2_f (s.e.)[g]		h^2_{gkin} (s.e.)[h]		%V_C[i]	logL[j]	n[k]
Height	0.51	(0.04)	0.23	(0.02)	0.51	(0.04)	74.59%	-20912.38	9,150
Weight	0.29	(0.06)	0.18	(0.03)	0.29	(0.06)	46.43%	10911.36	9,118
Fat	0.19	(0.06)	0.17	(0.03)	0.19	(0.06)	35.77%	-21502.48	8,926
BMI	0.20	(0.06)	0.20	(0.03)	0.20	(0.06)	40.03%	11514.74	9,107
Hips	0.19	(0.06)	0.15	(0.03)	0.19	(0.06)	34.05%	16791.17	8,984
Waist	0.17	(0.06)	0.18	(0.03)	0.17	(0.06)	34.92%	13523.39	9,016
WHR	0.20	(0.06)	0.07	(0.03)	0.20	(0.06)	27.28%	18148.01	8,995
ABSI	0.21	(0.06)	0.04	(0.03) [NS]	0.21	(0.06)	25.70%	19939.06	8,962
Urea	0.05	(0.06) [NS]	0.13	(0.03)	0.05	(0.06) [NS]	17.56%	8520.70	9,148
Creatinine	0.40	(0.05)	0.15	(0.03)	0.40	(0.05)	54.85%	12529.06	9,146
Glucose	0.09	(0.06) [NS]	0.05	(0.03)	0.09	(0.06) [NS]	13.98%	14808.47	8,936
TC	0.24	(0.06)	0.08	(0.03)	0.24	(0.06)	31.71%	10321.34	9,136
HDL	0.31	(0.06)	0.12	(0.03)	0.31	(0.06)	43.52%	8042.89	9,125
SBP	0.09	(0.06) [NS]	0.09	(0.03)	0.09	(0.06) [NS]	17.94%	14807.16	9,144
DBP	0.03	(0.06) [NS]	0.09	(0.03)	0.03	(0.06) [NS]	12.42%	14699.20	9,141
HR	0.13	(0.06)	0.10	(0.03)	0.13	(0.06)	23.03%	12058.00	9,126

[f] This column shows the proportion of total phenotypic variance captured by matrix **GRM_{kin}**

[g] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Family}**

[h] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[i] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[j] This column shows the log likelihood ratio for each analysis

[k] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (11) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'KS'

Trait	Model: KS[a]							
	σ^2_{kin} (s.e.)[b]		σ^2_{es} (s.e.)[c]		σ^2_{ε} (s.e.)[d]		V (s.e.)[e]	
Height	32.5294	(1.1390)	2.2467	(1.1089)	3.7308	(1.3523)	38.5069	(0.5950)
Weight	0.0189	(0.0012)	0.0015	(0.0013)[NS]	0.0135	(0.0015)	0.0339	(0.0005)
Fat	21.1965	(1.8067)	2.3099	(1.9175)[NS]	23.2642	(2.1961)	46.7706	(0.7133)
BMI	0.0151	(0.0011)	0.0012	(0.0011)[NS]	0.0132	(0.0013)	0.0295	(0.0004)
Hips	0.0038	(0.0003)	0.0002	(0.0003)[NS]	0.0047	(0.0004)	0.0087	(0.0001)
Waist	0.0083	(0.0007)	0.0004	(0.0007)[NS]	0.0096	(0.0009)	0.0183	(0.0003)
WHR	0.0020	(0.0003)	0.0000	(0.0003)[NS]	0.0044	(0.0003)	0.0064	(0.0001)
ABSI	0.0011	(0.0002)	0.0000	(0.0002)[NS]	0.0031	(0.0002)	0.0042	(0.0001)
Urea	0.0144	(0.0023)	0.0011	(0.0026)[NS]	0.0409	(0.0029)	0.0564	(0.0008)
Creatinine	0.0144	(0.0009)	0.0027	(0.0009)	0.0070	(0.0010)	0.0241	(0.0004)
Glucose	0.0020	(0.0006)	0.0012	(0.0007)[NS]	0.0098	(0.0007)	0.0131	(0.0002)
TC	0.0128	(0.0015)	0.0046	(0.0016)	0.0207	(0.0017)	0.0381	(0.0006)
HDL	0.0315	(0.0023)	0.0031	(0.0024)[NS]	0.0288	(0.0028)	0.0634	(0.0010)
SBP	0.0032	(0.0006)	0.0013	(0.0007)	0.0097	(0.0007)	0.0142	(0.0002)
DBP	0.0026	(0.0006)	0.0007	(0.0007)[NS]	0.0111	(0.0007)	0.0145	(0.0002)
HR	0.0077	(0.0010)	0.0008	(0.0012)[NS]	0.0174	(0.0013)	0.0258	(0.0004)

[a] Model '**KS**' = $\mathbf{GRM}_{kin}$ + $\mathbf{ERM}_{Sib}$

[b] This column shows the variance captured by matrix $\mathbf{GRM}_{kin}$

[c] This column shows the variance captured by matrix $\mathbf{ERM}_{Sib}$

[d] This column shows the residual variance

[e] This column shows the total phenotypic variance

Trait	Model: KS								
	h^2_{kin} (s.e.)[f]		e^2_s (s.e.)[g]		h^2_{gkin} (s.e.)[h]		%V_C[i]	logL[j]	n[k]
Height	0.84	(0.02)	0.06	(0.03)	0.84	(0.02)	90.31%	-20947.07	9,150
Weight	0.56	(0.03)	0.05	(0.04) [NS]	0.56	(0.03)	60.13%	10894.81	9,118
Fat	0.45	(0.04)	0.05	(0.04) [NS]	0.45	(0.04)	50.26%	-21516.18	8,926
BMI	0.51	(0.03)	0.04	(0.04) [NS]	0.51	(0.03)	55.32%	11494.03	9,107
Hips	0.43	(0.04)	0.02	(0.04) [NS]	0.43	(0.04)	45.62%	16779.31	8,984
Waist	0.45	(0.04)	0.02	(0.04) [NS]	0.45	(0.04)	47.65%	13507.57	9,016
WHR	0.32	(0.04)	0.00	(0.04) [NS]	0.32	(0.04)	31.61%	18145.38	8,995
ABSI	0.27	(0.04)	0.00	(0.05) [NS]	0.27	(0.04)	26.69%	19938.27	8,962
Urea	0.26	(0.04)	0.02	(0.05) [NS]	0.26	(0.04)	27.51%	8511.27	9,148
Creatinine	0.60	(0.03)	0.11	(0.04)	0.60	(0.03)	70.95%	12517.11	9,146
Glucose	0.16	(0.04)	0.09	(0.05) [NS]	0.16	(0.04)	24.80%	14807.96	8,936
TC	0.34	(0.04)	0.12	(0.04)	0.34	(0.04)	45.61%	10321.84	9,136
HDL	0.50	(0.03)	0.05	(0.04) [NS]	0.50	(0.03)	54.53%	8033.90	9,125
SBP	0.23	(0.04)	0.09	(0.05)	0.23	(0.04)	31.76%	14801.66	9,144
DBP	0.18	(0.04)	0.05	(0.05) [NS]	0.18	(0.04)	22.82%	14694.71	9,141
HR	0.30	(0.04)	0.03	(0.05) [NS]	0.30	(0.04)	32.76%	12051.22	9,126

[f] This column shows the proportion of total phenotypic variance captured by matrix **GRM_{kin}**

[g] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Sib}**

[h] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[i] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[j] This column shows the log likelihood ratio for each analysis

[k] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (12) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'KC'

Trait	Model: KC[a]							
	σ^2_{kin} (s.e.)[b]		σ^2_{ec} (s.e.)[c]		σ^2_{ε} (s.e.)[d]		V (s.e.)[e]	
Height	33.4731	(1.3932)	5.1655	(1.1006)	0.0001	(1.9381)^{NS}	38.6387	(0.5964)
Weight	0.0213	(0.0014)	0.0061	(0.0010)	0.0065	(0.0018)	0.0340	(0.0005)
Fat	24.3485	(1.9596)	8.7095	(1.4162)	13.8275	(2.5157)	46.8855	(0.7211)
BMI	0.0173	(0.0012)	0.0061	(0.0009)	0.0062	(0.0015)	0.0296	(0.0005)
Hips	0.0042	(0.0004)	0.0014	(0.0003)	0.0031	(0.0005)	0.0087	(0.0001)
Waist	0.0095	(0.0008)	0.0037	(0.0005)	0.0051	(0.0010)	0.0183	(0.0003)
WHR	0.0022	(0.0003)	0.0006	(0.0002)	0.0036	(0.0003)	0.0064	(0.0001)
ABSI	0.0012	(0.0002)	0.0002	(0.0001)	0.0028	(0.0002)	0.0042	(0.0001)
Urea	0.0162	(0.0023)	0.0073	(0.0017)	0.0331	(0.0030)	0.0565	(0.0008)
Creatinine	0.0169	(0.0010)	0.0039	(0.0007)	0.0034	(0.0012)	0.0242	(0.0004)
Glucose	0.0023	(0.0005)	0.0007	(0.0004)	0.0100	(0.0007)	0.0131	(0.0002)
TC	0.0148	(0.0015)	0.0030	(0.0011)	0.0203	(0.0020)	0.0381	(0.0006)
HDL	0.0353	(0.0025)	0.0092	(0.0018)	0.0191	(0.0033)	0.0635	(0.0010)
SBP	0.0038	(0.0006)	0.0015	(0.0004)	0.0089	(0.0007)	0.0142	(0.0002)
DBP	0.0030	(0.0006)	0.0013	(0.0004)	0.0102	(0.0007)	0.0145	(0.0002)
HR	0.0084	(0.0011)	0.0024	(0.0008)	0.0150	(0.0014)	0.0258	(0.0004)

[a] Model '**KC**' = $\mathbf{GRM}_{kin}$ + $\mathbf{ERM}_{couple}$

[b] This column shows the variance captured by matrix $\mathbf{GRM}_{kin}$

[c] This column shows the variance captured by matrix $\mathbf{ERM}_{Couple}$

[d] This column shows the residual variance

[e] This column shows the total phenotypic variance

Trait	Model: KC								
	h^2_{kin} (s.e.)[f]		e^2_c (s.e.)[g]		h^2_{gkin} (s.e.)[h]		%V_C[i]	logL[j]	n[k]
Height	0.87	(0.03)	0.13	(0.03)	0.87	(0.03)	100.00%	-20920.55	9,150
Weight	0.63	(0.04)	0.18	(0.03)	0.63	(0.04)	80.75%	10911.09	9,118
Fat	0.52	(0.04)	0.19	(0.03)	0.52	(0.04)	70.51%	-21500.82	8,926
BMI	0.58	(0.04)	0.21	(0.03)	0.58	(0.04)	79.15%	11515.40	9,107
Hips	0.48	(0.04)	0.16	(0.03)	0.48	(0.04)	64.65%	16791.83	8,984
Waist	0.52	(0.04)	0.20	(0.03)	0.52	(0.04)	72.02%	13527.40	9,016
WHR	0.34	(0.04)	0.09	(0.03)	0.34	(0.04)	43.41%	18149.31	8,995
ABSI	0.29	(0.04)	0.05	(0.03)	0.29	(0.04)	33.92%	19939.66	8,962
Urea	0.29	(0.04)	0.13	(0.03)	0.29	(0.04)	41.46%	8520.23	9,148
Creatinine	0.70	(0.04)	0.16	(0.03)	0.70	(0.04)	85.74%	12530.09	9,146
Glucose	0.18	(0.04)	0.05	(0.03)	0.18	(0.04)	23.15%	14808.62	8,936
TC	0.39	(0.04)	0.08	(0.03)	0.39	(0.04)	46.77%	10321.22	9,136
HDL	0.56	(0.04)	0.14	(0.03)	0.56	(0.04)	69.99%	8045.47	9,125
SBP	0.27	(0.04)	0.10	(0.03)	0.27	(0.04)	37.37%	14808.13	9,144
DBP	0.20	(0.04)	0.09	(0.03)	0.20	(0.04)	29.67%	14699.03	9,141
HR	0.33	(0.04)	0.09	(0.03)	0.33	(0.04)	41.88%	12056.17	9,126

[f] This column shows the proportion of total phenotypic variance captured by matrix **GRM_{kin}**

[g] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Couple}**

[h] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[i] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[j] This column shows the log likelihood ratio for each analysis

[k] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (13) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'FS'

Trait	Model: FS[a]							
	σ^2_{ef} (s.e.)[b]		σ^2_{es} (s.e.)[c]		σ^2_{ε} (s.e.)[d]		V (s.e.)[e]	
Height	15.9889	(0.7262)	0.6513	(1.3313)^{NS}	22.1549	(1.2259)	38.7952	(0.6066)
Weight	0.0094	(0.0007)	0.0005	(0.0014)^{NS}	0.0242	(0.0013)	0.0340	(0.0005)
Fat	10.9384	(0.9493)	1.6991	(1.9882)^{NS}	34.2604	(1.9081)	46.8979	(0.7206)
BMI	0.0080	(0.0006)	0.0002	(0.0012)^{NS}	0.0215	(0.0012)	0.0296	(0.0005)
Hips	0.0019	(0.0002)	0.0000	(0.0004)^{NS}	0.0068	(0.0004)	0.0087	(0.0001)
Waist	0.0043	(0.0004)	0.0000	(0.0008)^{NS}	0.0140	(0.0008)	0.0183	(0.0003)
WHR	0.0009	(0.0001)	0.0000	(0.0003)^{NS}	0.0055	(0.0003)	0.0064	(0.0001)
ABSI	0.0005	(0.0001)	0.0000	(0.0002)^{NS}	0.0037	(0.0002)	0.0042	(0.0001)
Urea	0.0080	(0.0011)	0.0006	(0.0026)^{NS}	0.0479	(0.0025)	0.0565	(0.0008)
Creatinine	0.0062	(0.0005)	0.0029	(0.0010)	0.0150	(0.0009)	0.0241	(0.0004)
Glucose	0.0009	(0.0002)	0.0014	(0.0007)	0.0108	(0.0007)	0.0131	(0.0002)
TC	0.0054	(0.0007)	0.0054	(0.0016)	0.0273	(0.0015)	0.0381	(0.0006)
HDL	0.0143	(0.0012)	0.0025	(0.0026)^{NS}	0.0467	(0.0025)	0.0635	(0.0010)
SBP	0.0016	(0.0003)	0.0012	(0.0007)^{NS}	0.0113	(0.0007)	0.0142	(0.0002)
DBP	0.0014	(0.0003)	0.0005	(0.0007)^{NS}	0.0126	(0.0006)	0.0145	(0.0002)
HR	0.0037	(0.0005)	0.0006	(0.0012)^{NS}	0.0216	(0.0012)	0.0258	(0.0004)

[a] Model '**FS**' = $\mathbf{ERM_{Family}}$ + $\mathbf{ERM_{Sib}}$

[b] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

[d] This column shows the residual variance

[e] This column shows the total phenotypic variance

Trait	Model: FS						
	e^2_f (s.e.)[f]		e^2_s (s.e.)[g]		%V_C[h]	logL[i]	n[j]
Height	0.41	(0.02)	0.02	(0.03) [NS]	42.89%	-20971.08	9,150
Weight	0.28	(0.02)	0.01	(0.04) [NS]	28.98%	10899.32	9,118
Fat	0.23	(0.02)	0.04	(0.04) [NS]	26.95%	-21506.57	8,926
BMI	0.27	(0.02)	0.01	(0.04) [NS]	27.57%	11509.15	9,107
Hips	0.22	(0.02)	0.00	(0.04) [NS]	21.98%	16786.97	8,984
Waist	0.24	(0.02)	0.00	(0.04) [NS]	23.59%	13519.62	9,016
WHR	0.14	(0.02)	0.00	(0.05) [NS]	14.36%	18143.32	8,995
ABSI	0.11	(0.02)	0.00	(0.05) [NS]	11.22%	19933.99	8,962
Urea	0.14	(0.02)	0.01	(0.05) [NS]	15.23%	8520.41	9,148
Creatinine	0.26	(0.02)	0.12	(0.04)	37.83%	12504.40	9,146
Glucose	0.07	(0.02)	0.10	(0.05)	17.26%	14808.80	8,936
TC	0.14	(0.02)	0.14	(0.04)	28.25%	10318.44	9,136
HDL	0.23	(0.02)	0.04	(0.04) [NS]	26.45%	8028.33	9,125
SBP	0.11	(0.02)	0.09	(0.05) [NS]	19.72%	14807.38	9,144
DBP	0.10	(0.02)	0.03	(0.05) [NS]	13.15%	14699.32	9,141
HR	0.14	(0.02)	0.02	(0.05) [NS]	16.70%	12056.06	9,126

[f] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Family}**

[g] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Sib}**

[h] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[i] This column shows the log likelihood ratio for each analysis

[j] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (14) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'FC'

Trait	Model: FC[a]							
	σ^2_{ef} (s.e.)[b]		σ^2_{ec} (s.e.)[c]		σ^2_{ε} (s.e.)[d]		V (s.e.)[e]	
Height	15.0512	(0.7917)	0.0001	(1.0230) [NS]	23.3193	(0.8291)	38.3706	(0.5951)
Weight	0.0087	(0.0007)	0.0000	(0.0011) [NS]	0.0252	(0.0009)	0.0339	(0.0005)
Fat	10.5649	(1.0195)	0.0001	(1.5363) [NS]	36.2925	(1.3932)	46.8575	(0.7176)
BMI	0.0076	(0.0006)	0.0000	(0.0009) [NS]	0.0220	(0.0008)	0.0296	(0.0005)
Hips	0.0018	(0.0002)	0.0000	(0.0003) [NS]	0.0069	(0.0003)	0.0087	(0.0001)
Waist	0.0042	(0.0004)	0.0000	(0.0006) [NS]	0.0141	(0.0005)	0.0183	(0.0003)
WHR	0.0008	(0.0001)	0.0000	(0.0002) [NS]	0.0055	(0.0002)	0.0064	(0.0001)
ABSI	0.0004	(0.0001)	0.0000	(0.0002) [NS]	0.0038	(0.0001)	0.0042	(0.0001)
Urea	0.0079	(0.0012)	0.0000	(0.0019) [NS]	0.0486	(0.0017)	0.0565	(0.0008)
Creatinine	0.0061	(0.0005)	0.0000	(0.0007) [NS]	0.0178	(0.0006)	0.0239	(0.0004)
Glucose	0.0009	(0.0003)	0.0000	(0.0005) [NS]	0.0122	(0.0004)	0.0131	(0.0002)
TC	0.0053	(0.0008)	0.0000	(0.0013) [NS]	0.0327	(0.0011)	0.0380	(0.0006)
HDL	0.0133	(0.0013)	0.0000	(0.0020) [NS]	0.0500	(0.0018)	0.0633	(0.0010)
SBP	0.0016	(0.0003)	0.0000	(0.0005) [NS]	0.0125	(0.0004)	0.0142	(0.0002)
DBP	0.0014	(0.0003)	0.0000	(0.0005) [NS]	0.0130	(0.0005)	0.0145	(0.0002)
HR	0.0034	(0.0005)	0.0000	(0.0009) [NS]	0.0224	(0.0008)	0.0258	(0.0004)

[a] Model '**FC**' = $\mathbf{ERM}_{\mathbf{Family}} + \mathbf{ERM}_{\mathbf{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{ERM}_{\mathbf{Family}}$

[c] This column shows the variance captured by matrix $\mathbf{ERM}_{\mathbf{Couple}}$

[d] This column shows the residual variance

[e] This column shows the total phenotypic variance

Trait	Model: FC						
	e^2_f (s.e.)[f]		e^2_c (s.e.)[g]		$\%V_C$[h]	logL[i]	n[j]
Height	0.39	(0.02)	0.00	(0.03) [NS]	39.23%	-20972.46	9,150
Weight	0.26	(0.02)	0.00	(0.03) [NS]	25.65%	10898.57	9,118
Fat	0.23	(0.02)	0.00	(0.03) [NS]	22.55%	-21507.16	8,926
BMI	0.26	(0.02)	0.00	(0.03) [NS]	25.60%	11508.84	9,107
Hips	0.21	(0.02)	0.00	(0.03) [NS]	20.63%	16786.70	8,984
Waist	0.23	(0.02)	0.00	(0.03) [NS]	22.93%	13519.51	9,016
WHR	0.13	(0.02)	0.00	(0.03) [NS]	13.20%	18143.02	8,995
ABSI	0.10	(0.02)	0.00	(0.04) [NS]	9.83%	19933.57	8,962
Urea	0.14	(0.02)	0.00	(0.03) [NS]	13.94%	8520.37	9,148
Creatinine	0.25	(0.02)	0.00	(0.03) [NS]	25.46%	12500.31	9,146
Glucose	0.07	(0.02)	0.00	(0.03) [NS]	6.87%	14807.36	8,936
TC	0.14	(0.02)	0.00	(0.03) [NS]	13.94%	10312.55	9,136
HDL	0.21	(0.02)	0.00	(0.03) [NS]	21.02%	8027.14	9,125
SBP	0.12	(0.02)	0.00	(0.03) [NS]	11.62%	14806.03	9,144
DBP	0.10	(0.02)	0.00	(0.03) [NS]	9.90%	14699.04	9,141
HR	0.13	(0.02)	0.00	(0.03) [NS]	13.24%	12055.68	9,126

[f] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Family}**

[g] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Couple}**

[h] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[i] This column shows the log likelihood ratio for each analysis

[j] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (15) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'SC'

Trait	Model: SC[a]							
	σ^2_{es} (s.e.)[b]		σ^2_{ec} (s.e.)[c]		σ^2_{ε} (s.e.)[d]		V (s.e.)[e]	
Height	19.3378	(1.2593)	10.4481	(1.0704)	9.0627	(1.5170)	38.8485	(0.5873)
Weight	0.0098	(0.0014)	0.0064	(0.0010)	0.0177	(0.0016)	0.0339	(0.0005)
Fat	12.5754	(1.9625)	8.9646	(1.4252)	25.2799	(2.3647)	46.8199	(0.7087)
BMI	0.0077	(0.0012)	0.0065	(0.0009)	0.0153	(0.0014)	0.0295	(0.0004)
Hips	0.0018	(0.0004)	0.0015	(0.0003)	0.0055	(0.0004)	0.0087	(0.0001)
Waist	0.0041	(0.0008)	0.0038	(0.0005)	0.0103	(0.0009)	0.0183	(0.0003)
WHR	0.0009	(0.0003)	0.0006	(0.0002)	0.0049	(0.0003)	0.0064	(0.0001)
ABSI	0.0005	(0.0002)	0.0002	(0.0001)	0.0035	(0.0002)	0.0042	(0.0001)
Urea	0.0086	(0.0025)	0.0073	(0.0017)	0.0406	(0.0030)	0.0564	(0.0008)
Creatinine	0.0099	(0.0009)	0.0032	(0.0007)	0.0109	(0.0011)	0.0241	(0.0004)
Glucose	0.0022	(0.0007)	0.0007	(0.0004)	0.0102	(0.0008)	0.0131	(0.0002)
TC	0.0109	(0.0015)	0.0026	(0.0011)	0.0245	(0.0019)	0.0381	(0.0006)
HDL	0.0185	(0.0025)	0.0092	(0.0018)	0.0358	(0.0030)	0.0635	(0.0009)
SBP	0.0028	(0.0007)	0.0014	(0.0004)	0.0099	(0.0008)	0.0142	(0.0002)
DBP	0.0018	(0.0006)	0.0013	(0.0004)	0.0113	(0.0008)	0.0145	(0.0002)
HR	0.0044	(0.0012)	0.0024	(0.0008)	0.0191	(0.0014)	0.0258	(0.0004)

[a] Model '**SC**' = $\mathbf{ERM}_{\mathbf{Sib}}$ + $\mathbf{ERM}_{\mathbf{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{ERM}_{\mathbf{Sib}}$

[c] This column shows the variance captured by matrix $\mathbf{ERM}_{\mathbf{Couple}}$

[d] This column shows the residual variance

[e] This column shows the total phenotypic variance

Trait	Model: SC						
	e^2_s (s.e.)[f]		e^2_c (s.e.)[g]		%V_C[h]	logL[i]	n[j]
Height	0.50	(0.03)	0.27	(0.03)	76.67%	-21172.72	9,150
Weight	0.29	(0.04)	0.19	(0.03)	47.71%	10815.32	9,118
Fat	0.27	(0.04)	0.19	(0.03)	46.01%	-21559.98	8,926
BMI	0.26	(0.04)	0.22	(0.03)	48.22%	11435.93	9,107
Hips	0.20	(0.04)	0.17	(0.03)	37.10%	16738.29	8,984
Waist	0.23	(0.04)	0.21	(0.03)	43.49%	13467.20	9,016
WHR	0.15	(0.04)	0.09	(0.03)	23.82%	18118.30	8,995
ABSI	0.11	(0.04)	0.06	(0.03)	17.06%	19916.86	8,962
Urea	0.15	(0.04)	0.13	(0.03)	28.13%	8500.85	9,148
Creatinine	0.41	(0.04)	0.14	(0.03)	54.49%	12428.60	9,146
Glucose	0.17	(0.05)	0.05	(0.03)	22.23%	14803.69	8,936
TC	0.29	(0.04)	0.07	(0.03)	35.57%	10292.71	9,136
HDL	0.29	(0.04)	0.14	(0.03)	43.65%	7965.86	9,125
SBP	0.20	(0.05)	0.10	(0.03)	29.88%	14794.53	9,144
DBP	0.13	(0.04)	0.09	(0.03)	21.84%	14689.23	9,141
HR	0.17	(0.05)	0.09	(0.03)	26.20%	12029.77	9,126

[f] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Sib}**

[g] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Couple}**

[h] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[i] This column shows the log likelihood ratio for each analysis

[j] This column shows the number of records used for each analysis

Table S2.4 (16) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GKF'

Trait	Model: GKF[a]							
	σ^2_g (s.e.)[b]		σ^2_{kin} (s.e.)[c]		σ^2_{ef} (s.e.)[d]		σ^2_ε (s.e.)[e]	
Height	20.7744	(1.5657)	0.0001	(2.1273)^{NS}	8.2288	(0.9656)	9.6680	(0.9908)
Weight	0.0095	(0.0014)	0.0004	(0.0023)^{NS}	0.0060	(0.0009)	0.0182	(0.0013)
Fat	11.2626	(1.9559)	0.0001	(3.3494)^{NS}	6.8154	(1.3862)	28.8660	(1.9075)
BMI	0.0070	(0.0012)	0.0000	(0.0020)^{NS}	0.0054	(0.0008)	0.0172	(0.0011)
Hips	0.0018	(0.0004)	0.0000	(0.0006)^{NS}	0.0013	(0.0003)	0.0057	(0.0004)
Waist	0.0030	(0.0007)	0.0001	(0.0013)^{NS}	0.0033	(0.0005)	0.0120	(0.0008)
WHR	0.0010	(0.0003)	0.0003	(0.0005)^{NS}	0.0005	(0.0002)	0.0046	(0.0003)
ABSI	0.0004	(0.0002)	0.0005	(0.0003)^{NS}	0.0002	(0.0001)^{NS}	0.0032	(0.0002)
Urea	0.0069	(0.0022)	0.0000	(0.0041)^{NS}	0.0051	(0.0016)	0.0446	(0.0025)
Creatinine	0.0060	(0.0010)	0.0034	(0.0015)	0.0039	(0.0006)	0.0110	(0.0008)
Glucose	0.0021	(0.0005)	0.0000	(0.0010)^{NS}	0.0002	(0.0004)^{NS}	0.0108	(0.0006)
TC	0.0057	(0.0015)	0.0035	(0.0027)^{NS}	0.0030	(0.0011)	0.0259	(0.0016)
HDL	0.0190	(0.0026)	0.0003	(0.0043)^{NS}	0.0081	(0.0017)	0.0361	(0.0024)
SBP	0.0018	(0.0006)	0.0000	(0.0010)^{NS}	0.0011	(0.0004)	0.0113	(0.0006)
DBP	0.0015	(0.0006)	0.0000	(0.0011)^{NS}	0.0008	(0.0004)	0.0122	(0.0006)
HR	0.0036	(0.0010)	0.0000	(0.0019)^{NS}	0.0025	(0.0007)	0.0197	(0.0011)

[a] Model '**GKF**' = $\mathbf{GRM_g}$ + $\mathbf{GRM_{kin}}$ + $\mathbf{ERM_{Family}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[e] This column shows the residual variance

Trait	Model: GKF							
	V (s.e.)[f]		h^2_g (s.e.)[g]		h^2_{kin} (s.e.)[h]		e^2_f (s.e.)[i]	
Height	38.6713	(0.6146)	0.54	(0.04)	0.00	(0.06) [NS]	0.21	(0.02)
Weight	0.0341	(0.0005)	0.28	(0.04)	0.01	(0.07) [NS]	0.18	(0.03)
Fat	46.9440	(0.7241)	0.24	(0.04)	0.00	(0.07) [NS]	0.15	(0.03)
BMI	0.0296	(0.0005)	0.24	(0.04)	0.00	(0.07) [NS]	0.18	(0.03)
Hips	0.0087	(0.0001)	0.20	(0.04)	0.00	(0.07) [NS]	0.15	(0.03)
Waist	0.0183	(0.0003)	0.16	(0.04)	0.00	(0.07) [NS]	0.18	(0.03)
WHR	0.0064	(0.0001)	0.15	(0.04)	0.05	(0.07) [NS]	0.07	(0.03)
ABSI	0.0042	(0.0001)	0.10	(0.04)	0.11	(0.07) [NS]	0.04	(0.03) [NS]
Urea	0.0565	(0.0008)	0.12	(0.04)	0.00	(0.07) [NS]	0.09	(0.03)
Creatinine	0.0242	(0.0004)	0.25	(0.04)	0.14	(0.06)	0.16	(0.03)
Glucose	0.0131	(0.0002)	0.16	(0.04)	0.00	(0.07) [NS]	0.01	(0.03) [NS]
TC	0.0381	(0.0006)	0.15	(0.04)	0.09	(0.07) [NS]	0.08	(0.03)
HDL	0.0636	(0.0010)	0.30	(0.04)	0.00	(0.07) [NS]	0.13	(0.03)
SBP	0.0142	(0.0002)	0.13	(0.04)	0.00	(0.07) [NS]	0.08	(0.03)
DBP	0.0145	(0.0002)	0.10	(0.04)	0.00	(0.07) [NS]	0.06	(0.03)
HR	0.0259	(0.0004)	0.14	(0.04)	0.00	(0.07) [NS]	0.10	(0.03)

[f] This column shows the total phenotypic variance

[g] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_g}$

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_{kin}}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Family}}$

Trait	Model: GKF				
	h^2_{gkin} (s.e.)[j]		%V_C[k]	logL[l]	n[m]
Height	0.54	(0.05)	75.00%	-20805.37	9,150
Weight	0.29	(0.06)	46.60%	10936.82	9,118
Fat	0.24	(0.06)	38.51%	-21481.62	8,926
BMI	0.24	(0.06)	41.89%	11535.99	9,107
Hips	0.20	(0.06)	35.00%	16805.37	8,984
Waist	0.16	(0.06)	34.78%	13531.90	9,016
WHR	0.20	(0.06)	26.93%	18155.22	8,995
ABSI	0.21	(0.06)	26.33%	19942.05	8,962
Urea	0.12	(0.06)	21.07%	8527.97	9,148
Creatinine	0.39	(0.05)	54.73%	12549.32	9,146
Glucose	0.16	(0.06)	17.58%	14819.82	8,936
TC	0.24	(0.06)	32.03%	10329.09	9,136
HDL	0.30	(0.06)	43.22%	8072.52	9,125
SBP	0.13	(0.06)	20.32%	14813.95	9,144
DBP	0.10	(0.06)	15.96%	14704.46	9,141
HR	0.14	(0.06)	23.85%	12064.94	9,126

[j] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[k] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[l] This column shows the log likelihood ratio for each analysis

[m] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (17) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GKS'

Trait	Model: GKS[a]							
	σ^2_g (s.e.)[b]		σ^2_{kin} (s.e.)[c]		σ^2_{es} (s.e.)[d]		σ^2_ε (s.e.)[e]	
Height	21.5688	(1.5907)	10.8293	(1.7402)	2.7072	(1.0569)	3.4257	(1.2776)
Weight	0.0095	(0.0014)	0.0096	(0.0018)	0.0015	(0.0012)^{NS}	0.0134	(0.0015)
Fat	12.3959	(1.9804)	9.0048	(2.5484)	2.247	(1.8987)^{NS}	23.2016	(2.1755)
BMI	0.0075	(0.0012)	0.0077	(0.0016)	0.0012	(0.0011)^{NS}	0.0132	(0.0013)
Hips	0.0018	(0.0004)	0.0020	(0.0005)	0.0002	(0.0003)^{NS}	0.0047	(0.0004)
Waist	0.0029	(0.0007)	0.0054	(0.0010)	0.0004	(0.0007)^{NS}	0.0095	(0.0009)
WHR	0.0009	(0.0003)	0.0011	(0.0004)	0.0000	(0.0003)^{NS}	0.0044	(0.0003)
ABSI	0.0004	(0.0002)	0.0008	(0.0002)	0.0000	(0.0002)^{NS}	0.0031	(0.0002)
Urea	0.0088	(0.0022)	0.0056	(0.0031)	0.0010	(0.0026)^{NS}	0.0410	(0.0029)
Creatinine	0.0056	(0.0010)	0.0088	(0.0013)	0.0026	(0.0009)	0.0070	(0.0010)
Glucose	0.0024	(0.0005)	0.0000	(0.0007)^{NS}	0.0010	(0.0007)^{NS}	0.0096	(0.0007)
TC	0.0058	(0.0015)	0.0070	(0.0021)	0.0049	(0.0016)	0.0204	(0.0017)
HDL	0.0187	(0.0026)	0.0128	(0.0033)	0.0028	(0.0024)^{NS}	0.0291	(0.0028)
SBP	0.0021	(0.0006)	0.0011	(0.0008)^{NS}	0.0014	(0.0007)	0.0095	(0.0007)
DBP	0.0020	(0.0006)	0.0006	(0.0008)^{NS}	0.0007	(0.0007)^{NS}	0.0111	(0.0007)
HR	0.0038	(0.0010)	0.0040	(0.0014)	0.0007	(0.0012)^{NS}	0.0174	(0.0013)

[a] Model '**GKS**' = $\mathbf{GRM_g} + \mathbf{GRM_{kin}} + \mathbf{ERM_{Sib}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

[e] This column shows the residual variance

Trait	Model: GKS							
	V (s.e.)[f]		h^2_g (s.e.)[g]		h^2_{kin} (s.e.)[h]		e^2_s (s.e.)[i]	
Height	38.5311	(0.6076)	0.56	(0.04)	0.28	(0.05)	0.07	(0.03)
Weight	0.0340	(0.0005)	0.28	(0.04)	0.28	(0.05)	0.05	(0.04) [NS]
Fat	46.8494	(0.7193)	0.26	(0.04)	0.19	(0.05)	0.05	(0.04) [NS]
BMI	0.0295	(0.0005)	0.25	(0.04)	0.26	(0.05)	0.04	(0.04) [NS]
Hips	0.0087	(0.0001)	0.21	(0.04)	0.23	(0.05)	0.02	(0.04) [NS]
Waist	0.0183	(0.0003)	0.16	(0.04)	0.30	(0.05)	0.02	(0.04) [NS]
WHR	0.0064	(0.0001)	0.15	(0.04)	0.17	(0.06)	0.00	(0.04) [NS]
ABSI	0.0042	(0.0001)	0.09	(0.04)	0.18	(0.06)	0.00	(0.05) [NS]
Urea	0.0565	(0.0008)	0.16	(0.04)	0.10	(0.05)	0.02	(0.05) [NS]
Creatinine	0.0241	(0.0004)	0.23	(0.04)	0.37	(0.05)	0.11	(0.04)
Glucose	0.0131	(0.0002)	0.19	(0.04)	0.00	(0.06) [NS]	0.08	(0.05) [NS]
TC	0.0381	(0.0006)	0.15	(0.04)	0.18	(0.05)	0.13	(0.04)
HDL	0.0635	(0.0010)	0.30	(0.04)	0.20	(0.05)	0.04	(0.04) [NS]
SBP	0.0142	(0.0002)	0.15	(0.04)	0.08	(0.06) [NS]	0.10	(0.05)
DBP	0.0145	(0.0002)	0.14	(0.04)	0.04	(0.05 [NS]	0.05	(0.05) [NS]
HR	0.0259	(0.0004)	0.15	(0.04)	0.15	(0.05)	0.03	(0.05) [NS]

[f] This column shows the total phenotypic variance

[g] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_g}$

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_{kin}}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Sib}}$

Trait	Model: GKS				
	h^2_{gkin} (s.e.)[j]		%V_C[k]	logL[l]	n[m]
Height	0.84	(0.05)	91.11%	-20834.93	9,150
Weight	0.56	(0.05)	60.48%	10919.79	9,118
Fat	0.46	(0.05)	50.48%	-21495.27	8,926
BMI	0.51	(0.05)	55.51%	11514.75	9,107
Hips	0.44	(0.05)	45.91%	16792.42	8,984
Waist	0.45	(0.05)	47.69%	13515.44	9,016
WHR	0.32	(0.05)	31.57%	18152.63	8,995
ABSI	0.27	(0.05)	26.51%	19941.16	8,962
Urea	0.26	(0.05)	27.28%	8519.69	9,148
Creatinine	0.60	(0.05)	70.89%	12535.15	9,146
Glucose	0.19	(0.05)	26.18%	14820.40	8,936
TC	0.34	(0.05)	46.51%	10329.97	9,136
HDL	0.50	(0.05)	54.16%	8062.34	9,125
SBP	0.23	(0.05)	32.51%	14809.08	9,144
DBP	0.18	(0.05)	22.91%	14701.17	9,141
HR	0.30	(0.05)	32.66%	12058.28	9,126

[j] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[k] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[l] This column shows the log likelihood ratio for each analysis

[m] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (18) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GKC'

Trait	Model: GKC[a]							
	σ^2_g (s.e.)[b]		σ^2_{kin} (s.e.)[c]		σ^2_{ec} (s.e.)[d]		σ^2_ε (s.e.)[e]	
Height	18.3186	(1.5799)	13.9405	(1.9834)	6.3464	(1.0705)	0.0001	(1.8882)
Weight	0.0095	(0.0014)	0.0120	(0.0018)	0.0061	(0.001)	0.0065	(0.0017)
Fat	12.3778	(1.9591)	12.1117	(2.6468)	8.7696	(1.4018)	13.7021	(2.4864)
BMI	0.0075	(0.0012)	0.0098	(0.0016)	0.0062	(0.0008)	0.0061	(0.0015)
Hips	0.0018	(0.0004)	0.0024	(0.0005)	0.0015	(0.0003)	0.0030	(0.0005)
Waist	0.0030	(0.0007)	0.0065	(0.0010)	0.0038	(0.0005)	0.0051	(0.001)
WHR	0.0010	(0.0003)	0.0012	(0.0004)	0.0006	(0.0002)	0.0036	(0.0003)
ABSI	0.0004	(0.0002)	0.0008	(0.0002)	0.0002	(0.0001)	0.0028	(0.0002)
Urea	0.0090	(0.0022)	0.0072	(0.0031)	0.0073	(0.0016)	0.0330	(0.003)
Creatinine	0.0059	(0.001)	0.0111	(0.0013)	0.0040	(0.0007)	0.0032	(0.0012)
Glucose	0.0025	(0.0005)	0.0000	(0.0007) [NS]	0.0006	(0.0004)	0.0100	(0.0007)
TC	0.0057	(0.0015)	0.0093	(0.0021)	0.0030	(0.0011)	0.0201	(0.002)
HDL	0.0190	(0.0026)	0.0164	(0.0034)	0.0094	(0.0018)	0.0188	(0.0032)
SBP	0.0021	(0.0006)	0.0018	(0.0008)	0.0015	(0.0004)	0.0088	(0.0007)
DBP	0.0020	(0.0006)	0.001	(0.0008) [NS]	0.0013	(0.0004)	0.0102	(0.0007)
HR	0.0038	(0.001)	0.0047	(0.0014)	0.0024	(0.0008)	0.0150	(0.0014)

[a] Model '**GKC**' = $\mathbf{GRM_g} + \mathbf{GRM_{kin}} + \mathbf{ERM_{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Couple}}$

[e] This column shows the residual variance

Trait	Model: GKC							
	V (s.e.)[f]		h^2_g (s.e.)[g]		h^2_{kin} (s.e.)[h]		e^2_c (s.e.)[i]	
Height	38.6055	(0.6024)	0.47	(0.04)	0.36	(0.05)	0.16	(0.03)
Weight	0.0340	(0.0005)	0.28	(0.04)	0.35	(0.05)	0.18	(0.03)
Fat	46.9613	(0.7268)	0.26	(0.04)	0.26	(0.06)	0.19	(0.03)
BMI	0.0296	(0.0005)	0.25	(0.04)	0.33	(0.05)	0.21	(0.03)
Hips	0.0087	(0.0001)	0.21	(0.04)	0.27	(0.06)	0.17	(0.03)
Waist	0.0183	(0.0003)	0.16	(0.04)	0.36	(0.06)	0.20	(0.03)
WHR	0.0064	(0.0001)	0.15	(0.04)	0.19	(0.06)	0.09	(0.03)
ABSI	0.0042	(0.0001)	0.10	(0.04)	0.19	(0.06)	0.05	(0.03)
Urea	0.0565	(0.0008)	0.16	(0.04)	0.13	(0.06)	0.13	(0.03)
Creatinine	0.0242	(0.0004)	0.24	(0.04)	0.46	(0.05)	0.17	(0.03)
Glucose	0.0131	(0.0002)	0.19	(0.04)	0.00	(0.06) [NS]	0.05	(0.03)
TC	0.0381	(0.0006)	0.15	(0.04)	0.24	(0.05)	0.08	(0.03)
HDL	0.0636	(0.001)	0.30	(0.04)	0.26	(0.05)	0.15	(0.03)
SBP	0.0142	(0.0002)	0.15	(0.04)	0.13	(0.06)	0.10	(0.03)
DBP	0.0145	(0.0002)	0.14	(0.04)	0.07	(0.05) [NS]	0.09	(0.03)
HR	0.0259	(0.0004)	0.15	(0.04)	0.18	(0.06)	0.09	(0.03)

[f] This column shows the total phenotypic variance

[g] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_g}$

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_{kin}}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Couple}}$

Trait	Model: GKC				
	h^2_{gkin} (s.e.)[j]		%V_C[k]	logL[l]	n[m]
Height	0.84	(0.05)	100.00%	-20814.46	9,150
Weight	0.63	(0.05)	81.04%	10936.62	9,118
Fat	0.52	(0.05)	70.82%	-21479.41	8,926
BMI	0.59	(0.05)	79.56%	11537.17	9,107
Hips	0.49	(0.05)	65.51%	16806.12	8,984
Waist	0.52	(0.05)	72.18%	13535.79	9,016
WHR	0.34	(0.05)	43.31%	18156.42	8,995
ABSI	0.29	(0.05)	34.17%	19942.62	8,962
Urea	0.29	(0.05)	41.53%	8528.94	9,148
Creatinine	0.70	(0.05)	86.84%	12549.93	9,146
Glucose	0.19	(0.05)	23.95%	14821.27	8,936
TC	0.39	(0.05)	47.19%	10328.88	9,136
HDL	0.56	(0.05)	70.51%	8075.13	9,125
SBP	0.27	(0.05)	37.68%	14815.30	9,144
DBP	0.21	(0.05)	29.97%	14705.45	9,141
HR	0.33	(0.05)	42.03%	12063.24	9,126

[k] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[l] This column shows the log likelihood ratio for each analysis

[m] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (19) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GFS'

Trait	Model: GFS[a]							
	σ^2_g (s.e.)[b]		σ^2_{ef} (s.e.)[c]		σ^2_{es} (s.e.)[d]		σ^2_ε (s.e.)[e]	
Height	20.8692	(1.2479)	8.1139	(0.8092)	0.6718	(1.0319)^{NS}	9.0281	(1.2010)
Weight	0.0096	(0.0012)	0.0061	(0.0008)	0.0002	(0.0012)^{NS}	0.0182	(0.0014)
Fat	11.2889	(1.6421)	6.9576	(1.0844)	0.8331	(1.8754)^{NS}	27.8577	(2.0336)
BMI	0.0070	(0.0010)	0.0055	(0.0007)	0.0000	(0.0011)^{NS}	0.0171	(0.0012)
Hips	0.0017	(0.0003)	0.0013	(0.0002)	0.0000	(0.0003)^{NS}	0.0057	(0.0004)
Waist	0.0029	(0.0006)	0.0033	(0.0004)	0.0000	(0.0007)^{NS}	0.0121	(0.0008)
WHR	0.0010	(0.0002)	0.0005	(0.0001)	0.0000	(0.0003)^{NS}	0.0049	(0.0003)
ABSI	0.0005	(0.0001)	0.0003	(0.0001)	0.0000	(0.0002)^{NS}	0.0034	(0.0002)
Urea	0.0071	(0.0019)	0.0055	(0.0012)	0.0000	(0.0026)^{NS}	0.0440	(0.0027)
Creatinine	0.0072	(0.0008)	0.0045	(0.0005)	0.0017	(0.0009)	0.0109	(0.0009)
Glucose	0.0022	(0.0005)	0.0003	(0.0003)^{NS}	0.0010	(0.0007)^{NS}	0.0097	(0.0007)
TC	0.0066	(0.0013)	0.0034	(0.0008)	0.0047	(0.0016)	0.0235	(0.0016)
HDL	0.0191	(0.0022)	0.0081	(0.0014)	0.0008	(0.0024)^{NS}	0.0355	(0.0026)
SBP	0.0018	(0.0005)	0.0011	(0.0003)	0.0011	(0.0007)^{NS}	0.0102	(0.0007)
DBP	0.0016	(0.0005)	0.0009	(0.0003)	0.0004	(0.0007)^{NS}	0.0116	(0.0007)
HR	0.0036	(0.0009)	0.0026	(0.0006)	0.0002	(0.0012)^{NS}	0.0195	(0.0013)

[a] Model '**GFS**' = $\mathbf{GRM_g}$ + $\mathbf{ERM_{Family}}$ + $\mathbf{ERM_{Sib}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

[e] This column shows the residual variance

Trait	Model: GFS							
	V (s.e.)[f]		h^2_g (s.e.)[g]		e^2_f (s.e.)[h]		e^2_s (s.e.)[i]	
Height	38.6831	(0.6126)	0.54	(0.03)	0.21	(0.02)	0.02	(0.03) [NS]
Weight	0.0341	(0.0005)	0.28	(0.03)	0.18	(0.02)	0.00	(0.04) [NS]
Fat	46.9374	(0.7248)	0.24	(0.03)	0.15	(0.02)	0.02	(0.04) [NS]
BMI	0.0296	(0.0005)	0.24	(0.03)	0.19	(0.02)	0.00	(0.04) [NS]
Hips	0.0087	(0.0001)	0.20	(0.03)	0.15	(0.02)	0.00	(0.04) [NS]
Waist	0.0183	(0.0003)	0.16	(0.03)	0.18	(0.02)	0.00	(0.04) [NS]
WHR	0.0064	(0.0001)	0.16	(0.03)	0.08	(0.02)	0.00	(0.04) [NS]
ABSI	0.0042	(0.0001)	0.12	(0.03)	0.07	(0.02)	0.00	(0.05) [NS]
Urea	0.0565	(0.0008)	0.13	(0.03)	0.10	(0.02)	0.00	(0.05) [NS]
Creatinine	0.0242	(0.0004)	0.30	(0.03)	0.19	(0.02)	0.07	(0.04)
Glucose	0.0131	(0.0002)	0.16	(0.03)	0.02	(0.02) [NS]	0.08	(0.05) [NS]
TC	0.0381	(0.0006)	0.17	(0.03)	0.09	(0.02)	0.12	(0.04)
HDL	0.0636	(0.001)	0.30	(0.03)	0.13	(0.02)	0.01	(0.04) [NS]
SBP	0.0142	(0.0002)	0.13	(0.03)	0.08	(0.02)	0.08	(0.05) [NS]
DBP	0.0145	(0.0002)	0.11	(0.03)	0.06	(0.02)	0.03	(0.05) [NS]
HR	0.0259	(0.0004)	0.14	(0.03)	0.10	(0.02)	0.01	(0.05) [NS]

[f] This column shows the total phenotypic variance

[g] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_g}$

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Family}}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Sib}}$

Trait	Model: GFS				
	h^2_{gkin} (s.e.)[j]		%V_C[k]	logL[l]	n[m]
Height	0.54	(0.03)	76.66%	-20805.17	9,150
Weight	0.28	(0.03)	46.65%	10936.82	9,118
Fat	0.24	(0.03)	40.65%	-21481.47	8,926
BMI	0.24	(0.03)	42.18%	11536.01	9,107
Hips	0.20	(0.03)	34.75%	16805.36	8,984
Waist	0.16	(0.03)	33.81%	13531.86	9,016
WHR	0.16	(0.03)	23.91%	18154.85	8,995
ABSI	0.12	(0.03)	18.66%	19940.81	8,962
Urea	0.13	(0.03)	22.23%	8528.16	9,148
Creatinine	0.30	(0.03)	55.22%	12548.43	9,146
Glucose	0.16	(0.03)	26.34%	14820.89	8,936
TC	0.17	(0.03)	38.41%	10332.46	9,136
HDL	0.30	(0.03)	44.09%	8072.58	9,125
SBP	0.13	(0.03)	28.21%	14815.09	9,144
DBP	0.11	(0.03)	19.71%	14704.79	9,141
HR	0.14	(0.03)	24.76%	12064.96	9,126

[j] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[k] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[l] This column shows the log likelihood ratio for each analysis

[m] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (20) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GFC'

Trait	Model: GFC[a]							
	σ^2_g (s.e.)[b]		σ^2_{ef} (s.e.)[c]		σ^2_{ec} (s.e.)[d]		σ^2_ε (s.e.)[e]	
Height	21.7673	(1.5342)	7.7046	(0.9183)	1.2726	(1.0644)^{NS}	7.9914	(1.5949)
Weight	0.0098	(0.0014)	0.0060	(0.0009)	0.0003	(0.0012)^{NS}	0.0180	(0.0016)
Fat	12.9167	(1.9161)	5.8006	(1.3076)	3.0399	(1.7094)	25.2163	(2.2614)
BMI	0.0078	(0.0012)	0.0049	(0.0008)	0.0014	(0.0010)^{NS}	0.0155	(0.0014)
Hips	0.0019	(0.0003)	0.0012	(0.0002)	0.0003	(0.0003)^{NS}	0.0053	(0.0004)
Waist	0.0034	(0.0007)	0.0029	(0.0005)	0.0008	(0.0007)^{NS}	0.0111	(0.0009)
WHR	0.0011	(0.0003)	0.0005	(0.0002)	0.0000	(0.0002)^{NS}	0.0047	(0.0003)
ABSI	0.0004	(0.0002)	0.0003	(0.0001)	0.0000	(0.0002)^{NS}	0.0034	(0.0002)
Urea	0.0088	(0.0022)	0.0040	(0.0016)	0.0035	(0.0021)^{NS}	0.0403	(0.0027)
Creatinine	0.0059	(0.0009)	0.0049	(0.0007)	0.0000	(0.0008)^{NS}	0.0134	(0.0011)
Glucose	0.0025	(0.0005)	0.0000	(0.0004)^{NS}	0.0007	(0.0005)^{NS}	0.0098	(0.0006)
TC	0.0054	(0.0015)	0.0041	(0.0010)	0.0000	(0.0014)^{NS}	0.0286	(0.0018)
HDL	0.0204	(0.0025)	0.0073	(0.0017)	0.0020	(0.0022)^{NS}	0.0339	(0.0030)
SBP	0.0022	(0.0006)	0.0008	(0.0004)	0.0006	(0.0005)^{NS}	0.0105	(0.0007)
DBP	0.0020	(0.0006)	0.0006	(0.0004)^{NS}	0.0008	(0.0006)^{NS}	0.0112	(0.0007)
HR	0.0035	(0.0010)	0.0026	(0.0007)	0.0000	(0.0010)^{NS}	0.0198	(0.0013)

[a] Model '**GFC**' = $\mathbf{GRM_g}$ + $\mathbf{ERM_{Family}}$ + $\mathbf{ERM_{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Couple}}$

[e] This column shows the residual variance

Trait	Model: GFC							
	V (s.e.)[f]		h^2_g (s.e.)[g]		e^2_f (s.e.)[h]		e^2_c (s.e.)[i]	
Height	38.7359	(0.6176)	0.56	(0.04)	0.20	(0.02)	0.03	(0.03) [NS]
Weight	0.0341	(0.0005)	0.29	(0.04)	0.18	(0.03)	0.01	(0.03) [NS]
Fat	46.9735	(0.7270)	0.27	(0.04)	0.12	(0.03)	0.06	(0.04)
BMI	0.0296	(0.0005)	0.26	(0.04)	0.16	(0.03)	0.05	(0.03) [NS]
Hips	0.0087	(0.0001)	0.22	(0.04)	0.14	(0.03)	0.04	(0.04) [NS]
Waist	0.0183	(0.0003)	0.19	(0.04)	0.16	(0.03)	0.04	(0.04) [NS]
WHR	0.0064	(0.0001)	0.17	(0.04)	0.08	(0.03)	0.01	(0.04) [NS]
ABSI	0.0042	(0.0001)	0.11	(0.04)	0.07	(0.03)	0.00	(0.04) [NS]
Urea	0.0565	(0.0008)	0.16	(0.04)	0.07	(0.03)	0.06	(0.04) [NS]
Creatinine	0.0242	(0.0004)	0.24	(0.04)	0.20	(0.03)	0.00	(0.03) [NS]
Glucose	0.0131	(0.0002)	0.19	(0.04)	0.00	(0.03) [NS]	0.06	(0.04) [NS]
TC	0.0381	(0.0006)	0.14	(0.04)	0.11	(0.03)	0.00	(0.04) [NS]
HDL	0.0636	(0.0010)	0.32	(0.04)	0.11	(0.03)	0.03	(0.03) [NS]
SBP	0.0142	(0.0002)	0.15	(0.04)	0.06	(0.03)	0.05	(0.04) [NS]
DBP	0.0145	(0.0002)	0.14	(0.04)	0.04	(0.03) [NS]	0.05	(0.04) [NS]
HR	0.0259	(0.0004)	0.13	(0.04)	0.10	(0.03)	0.00	(0.04) [NS]

[f] This column shows the total phenotypic variance

[g] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_g}$

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Family}}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Couple}}$

Trait	Model: GFC				
	h^2_{gkin} (s.e.)[j]		%V_C[k]	logL[l]	n[m]
Height	0.56	(0.04)	79.37%	-20804.67	9,150
Weight	0.29	(0.04)	47.12%	10936.84	9,118
Fat	0.27	(0.04)	46.32%	-21480.09	8,926
BMI	0.26	(0.04)	47.53%	11536.89	9,107
Hips	0.22	(0.04)	39.03%	16805.80	8,984
Waist	0.19	(0.04)	39.14%	13532.58	9,016
WHR	0.17	(0.04)	25.20%	18154.98	8,995
ABSI	0.11	(0.04)	17.97%	19940.69	8,962
Urea	0.16	(0.04)	28.84%	8529.45	9,148
Creatinine	0.24	(0.04)	44.60%	12545.13	9,146
Glucose	0.19	(0.04)	24.83%	14821.27	8,936
TC	0.14	(0.04)	24.96%	10327.56	9,136
HDL	0.32	(0.04)	46.61%	8072.93	9,125
SBP	0.15	(0.04)	25.44%	14814.84	9,144
DBP	0.14	(0.04)	23.30%	14705.64	9,141
HR	0.13	(0.04)	23.48%	12064.92	9,126

[j] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[k] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[l] This column shows the log likelihood ratio for each analysis

[m] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (21) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GSC'

Trait	Model: GSC[a]							
	σ^2_g (s.e.)[b]		σ^2_{es} (s.e.)[c]		σ^2_{ec} (s.e.)[d]		σ^2_ε (s.e.)[e]	
Height	30.4130	(1.2064)	2.6343	(1.1799)	5.9003	(0.9684)	0.0001	(1.6301)
Weight	0.0153	(0.0011)	0.0021	(0.0013)^{NS}	0.0051	(0.0010)	0.0115	(0.0017)
Fat	17.9125	(1.5325)	3.3371	(1.9597)^{NS}	7.7790	(1.3970)	17.9231	(2.4792)
BMI	0.0121	(0.0010)	0.0017	(0.0012)^{NS}	0.0055	(0.0008)	0.0104	(0.0015)
Hips	0.0030	(0.0003)	0.0003	(0.0004)^{NS}	0.0013	(0.0003)	0.0041	(0.0005)
Waist	0.0059	(0.0006)	0.0011	(0.0008)^{NS}	0.0034	(0.0005)	0.0079	(0.0010)
WHR	0.0016	(0.0002)	0.0001	(0.0003)^{NS}	0.0005	(0.0002)	0.0042	(0.0004)
ABSI	0.0008	(0.0001)	0.0001	(0.0002)^{NS}	0.0002	(0.0001)^{NS}	0.0031	(0.0002)
Urea	0.0120	(0.0017)	0.0021	(0.0026)^{NS}	0.0070	(0.0016)	0.0353	(0.0032)
Creatinine	0.0107	(0.0008)	0.0039	(0.0009)	0.0028	(0.0007)	0.0067	(0.0012)
Glucose	0.0024	(0.0004)	0.0011	(0.0007)^{NS}	0.0007	(0.0004)	0.0089	(0.0008)
TC	0.0093	(0.0012)	0.0065	(0.0016)	0.0022	(0.0011)	0.0201	(0.0020)
HDL	0.0271	(0.0020)	0.0036	(0.0025)^{NS}	0.0081	(0.0018)	0.0249	(0.0032)
SBP	0.0027	(0.0004)	0.0016	(0.0007)	0.0014	(0.0004)	0.0085	(0.0008)
DBP	0.0024	(0.0004)	0.0007	(0.0006)^{NS}	0.0013	(0.0004)	0.0101	(0.0008)
HR	0.0059	(0.0008)	0.0012	(0.0012)^{NS}	0.0021	(0.0007)	0.0166	(0.0015)

[a] Model '**GSC**' = $\mathbf{GRM_g}$ + $\mathbf{ERM_{Sib}}$ + $\mathbf{ERM_{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Couple}}$

[e] This column shows the residual variance

Trait	Model: GSC							
	V (s.e.)[f]		h^2_g (s.e.)[g]		e^2_s (s.e.)[h]		e^2_c (s.e.)[i]	
Height	38.9477	(0.6254)	0.78	(0.02)	0.07	(0.03)	0.15	(0.02)
Weight	0.0340	(0.0005)	0.45	(0.03)	0.06	(0.04) [NS]	0.15	(0.03)
Fat	46.9517	(0.7252)	0.38	(0.03)	0.07	(0.04) [NS]	0.17	(0.03)
BMI	0.0296	(0.0005)	0.41	(0.03)	0.06	(0.04) [NS]	0.18	(0.03)
Hips	0.0087	(0.0001)	0.34	(0.03)	0.04	(0.04) [NS]	0.15	(0.03)
Waist	0.0183	(0.0003)	0.32	(0.03)	0.06	(0.04) [NS]	0.19	(0.03)
WHR	0.0064	(0.0001)	0.24	(0.03)	0.02	(0.04) [NS]	0.08	(0.03)
ABSI	0.0042	(0.0001)	0.19	(0.03)	0.01	(0.05) [NS]	0.05	(0.03) [NS]
Urea	0.0565	(0.0008)	0.21	(0.03)	0.04	(0.05) [NS]	0.12	(0.03)
Creatinine	0.0242	(0.0004)	0.44	(0.03)	0.16	(0.04)	0.12	(0.03)
Glucose	0.0131	(0.0002)	0.18	(0.03)	0.08	(0.05) [NS]	0.05	(0.03)
TC	0.0381	(0.0006)	0.24	(0.03)	0.17	(0.04)	0.06	(0.03)
HDL	0.0636	(0.0010)	0.43	(0.03)	0.06	(0.04) [NS]	0.13	(0.03)
SBP	0.0142	(0.0002)	0.19	(0.03)	0.11	(0.05)	0.10	(0.03)
DBP	0.0145	(0.0002)	0.16	(0.03)	0.05	(0.04) [NS]	0.09	(0.03)
HR	0.0259	(0.0004)	0.23	(0.03)	0.05	(0.05) [NS]	0.08	(0.03)

[f] This column shows the total phenotypic variance

[g] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_g}$

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Sib}}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Couple}}$

Trait	Model: GSC				
	h^2_{gkin} (s.e.)[j]		%V_C[k]	logL[l]	n[m]
Height	0.78	(0.02)	100.00%	-20837.64	9,150
Weight	0.45	(0.03)	66.00%	10916.80	9,118
Fat	0.38	(0.03)	61.83%	-21488.46	8,926
BMI	0.41	(0.03)	64.99%	11519.93	9,107
Hips	0.34	(0.03)	52.67%	16794.83	8,984
Waist	0.32	(0.03)	56.75%	13517.62	9,016
WHR	0.24	(0.03)	33.81%	18150.54	8,995
ABSI	0.19	(0.03)	26.26%	19936.67	8,962
Urea	0.21	(0.03)	37.40%	8526.67	9,148
Creatinine	0.44	(0.03)	72.25%	12520.46	9,146
Glucose	0.18	(0.03)	32.06%	14822.27	8,936
TC	0.24	(0.03)	47.32%	10326.26	9,136
HDL	0.43	(0.03)	60.86%	8064.52	9,125
SBP	0.19	(0.03)	40.09%	14815.00	9,144
DBP	0.16	(0.03)	30.12%	14705.25	9,141
HR	0.23	(0.03)	35.77%	12058.42	9,126

[j] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[k] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[l] This column shows the log likelihood ratio for each analysis

[m] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (22) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'KFS'

Trait	Model: KFS[a]							
	σ^2_{kin} (s.e.)[b]		σ^2_{ef} (s.e.)[c]		σ^2_{es} (s.e.)[d]		σ^2_{ε} (s.e.)[e]	
Height	19.8282	(1.7668)	9.0590	(1.0297)	0.0001	(1.0758)^{NS}	9.8702	(1.4985)
Weight	0.0098	(0.0019)	0.0060	(0.0010)	0.0002	(0.0013)^{NS}	0.0181	(0.0017)
Fat	8.6996	(2.8479)	7.8799	(1.3933)	1.0075	(1.9186)^{NS}	29.2975	(2.4709)
BMI	0.0060	(0.0017)	0.0059	(0.0008)	0.0000	(0.0011)^{NS}	0.0178	(0.0015)
Hips	0.0016	(0.0005)	0.0013	(0.0003)	0.0000	(0.0003)^{NS}	0.0058	(0.0005)
Waist	0.0031	(0.0011)	0.0032	(0.0005)	0.0000	(0.0007)^{NS}	0.0120	(0.0010)
WHR	0.0013	(0.0004)	0.0004	(0.0002)	0.0000	(0.0003)^{NS}	0.0047	(0.0004)
ABSI	0.0009	(0.0003)	0.0001	(0.0001)^{NS}	0.0000	(0.0002)^{NS}	0.0032	(0.0002)
Urea	0.0027	(0.0035)^{NS}	0.0071	(0.0016)	0.0002	(0.0026)^{NS}	0.0464	(0.0032)
Creatinine	0.0096	(0.0013)	0.0035	(0.0007)	0.0017	(0.0009)	0.0093	(0.0011)
Glucose	0.0010	(0.0008)^{NS}	0.0006	(0.0004)	0.0012	(0.0007)^{NS}	0.0103	(0.0008)
TC	0.0085	(0.0023)	0.0027	(0.0011)	0.0042	(0.0016)	0.0227	(0.0020)
HDL	0.0198	(0.0036)	0.0077	(0.0018)	0.0012	(0.0024)^{NS}	0.0348	(0.0032)
SBP	0.0011	(0.0009)^{NS}	0.0013	(0.0004)	0.0011	(0.0007)^{NS}	0.0107	(0.0008)
DBP	0.0004	(0.0009)^{NS}	0.0013	(0.0004)	0.0004	(0.0007)^{NS}	0.0123	(0.0008)
HR	0.0032	(0.0016)	0.0027	(0.0007)	0.0003	(0.0012)^{NS}	0.0197	(0.0015)

[a] Model '**KFS**' = $\mathbf{GRM_{kin}}$ + $\mathbf{ERM_{Family}}$ + $\mathbf{ERM_{Sib}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

[e] This column shows the residual variance

Trait	Model: KFS							
	V (s.e.)[f]		h^2_{kin} (s.e.)[g]		e^2_f (s.e.)[h]		e^2_s (s.e.)[i]	
Height	38.7575	(0.6073)	0.51	(0.04)	0.23	(0.03)	0.00	(0.03) NS
Weight	0.0340	(0.0005)	0.29	(0.06)	0.18	(0.03)	0.00	(0.04) NS
Fat	46.8845	(0.7205)	0.19	(0.06)	0.17	(0.03)	0.02	(0.04) NS
BMI	0.0296	(0.0005)	0.20	(0.06)	0.20	(0.03)	0.00	(0.04) NS
Hips	0.0087	(0.0001)	0.18	(0.06)	0.15	(0.03)	0.00	(0.04) NS
Waist	0.0183	(0.0003)	0.17	(0.06)	0.17	(0.03)	0.00	(0.04) NS
WHR	0.0064	(0.0001)	0.20	(0.06)	0.06	(0.03)	0.00	(0.04) NS
ABSI	0.0042	(0.0001)	0.21	(0.07)	0.03	(0.03) NS	0.00	(0.05) NS
Urea	0.0565	(0.0008)	0.05	(0.06) NS	0.13	(0.03)	0.00	(0.05) NS
Creatinine	0.0242	(0.0004)	0.40	(0.05)	0.14	(0.03)	0.07	(0.04)
Glucose	0.0131	(0.0002)	0.08	(0.06) NS	0.05	(0.03)	0.09	(0.05) NS
TC	0.0381	(0.0006)	0.22	(0.06)	0.07	(0.03)	0.11	(0.04)
HDL	0.0635	(0.001)	0.31	(0.06)	0.12	(0.03)	0.02	(0.04) NS
SBP	0.0142	(0.0002)	0.08	(0.06) NS	0.09	(0.03)	0.07	(0.05) NS
DBP	0.0145	(0.0002)	0.03	(0.06) NS	0.09	(0.03)	0.03	(0.05) NS
HR	0.0259	(0.0004)	0.12	(0.06)	0.10	(0.03)	0.01	(0.05) NS

[f] This column shows the total phenotypic variance

[g] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_{kin}}$

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Family}}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Sib}}$

Trait	Model: KFS				
		h^2_{gkin} (s.e.)[j]	$\%V_C$[k]	logL[l]	n[m]
Height	0.51	(0.04)	74.53%	-20912.38	9,150
Weight	0.29	(0.06)	46.98%	10911.37	9,118
Fat	0.19	(0.06)	37.51%	-21502.34	8,926
BMI	0.20	(0.06)	39.90%	11514.74	9,107
Hips	0.18	(0.06)	33.72%	16791.17	8,984
Waist	0.17	(0.06)	34.24%	13523.36	9,016
WHR	0.20	(0.06)	26.02%	18147.84	8,995
ABSI	0.21	(0.07)	23.02%	19938.80	8,962
Urea	0.05	(0.06) [NS]	17.72%	8520.70	9,148
Creatinine	0.40	(0.05)	61.41%	12530.88	9,146
Glucose	0.08	(0.06) [NS]	21.48%	14809.53	8,936
TC	0.22	(0.06)	40.30%	10324.72	9,136
HDL	0.31	(0.06)	45.17%	8043.02	9,125
SBP	0.08	(0.06) [NS]	24.67%	14808.17	9,144
DBP	0.03	(0.06) [NS]	14.25%	14699.43	9,141
HR	0.12	(0.06)	24.02%	12058.02	9,126

[j] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[k] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[l] This column shows the log likelihood ratio for each analysis

[m] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (23) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'KFC'

Trait	Model: KFC[a]							
	σ^2_{kin} (s.e.)[b]		σ^2_{ef} (s.e.)[c]		σ^2_{ec} (s.e.)[d]		σ^2_{ε} (s.e.)[e]	
Height	34.4389	(6.9369)	0.7251	(3.4886)^{NS}	6.4988	(3.5885)^{NS}	0.0001	(6.9818)^{NS}
Weight	0.0148	(0.0058)	0.0034	(0.0029)^{NS}	0.0028	(0.0030)^{NS}	0.0130	(0.0058)
Fat	22.4021	(8.1087)	1.0184	(4.1062)^{NS}	7.7209	(4.2528)	15.7462	(8.1801)^{NS}
BMI	0.0129	(0.0051)	0.0023	(0.0026)^{NS}	0.0039	(0.0027)^{NS}	0.0105	(0.0051)
Hips	0.0034	(0.0016)	0.0004	(0.0008)^{NS}	0.0010	(0.0008)^{NS}	0.0038	(0.0016)
Waist	0.0112	(0.0031)	0.0000	(0.0016)^{NS}	0.0044	(0.0016)^{NS}	0.0027	(0.0031)^{NS}
WHR	0.0028	(0.0011)	0.0000	(0.0006)^{NS}	0.0009	(0.0006)^{NS}	0.0027	(0.0011)
ABSI	0.0016	(0.0007)	0.0000	(0.0004)^{NS}	0.0004	(0.0004)^{NS}	0.0023	(0.0007)
Urea	0.0067	(0.0095)^{NS}	0.0051	(0.0049)^{NS}	0.0023	(0.0051)^{NS}	0.0425	(0.0097)
Creatinine	0.0156	(0.0042)	0.0006	(0.0021)^{NS}	0.0032	(0.0022)^{NS}	0.0047	(0.0042)^{NS}
Glucose	0.0022	(0.0022)^{NS}	0.0001	(0.0011)^{NS}	0.0006	(0.0012)^{NS}	0.0102	(0.0022)
TC	0.0111	(0.0067)	0.0020	(0.0034)^{NS}	0.0011	(0.0035)^{NS}	0.0240	(0.0068)
HDL	0.0410	(0.0103)	0.0000	(0.0052)^{NS}	0.0114	(0.0054)^{NS}	0.0109	(0.0104)^{NS}
SBP	0.0046	(0.0025)^{NS}	0.0000	(0.0013)^{NS}	0.0017	(0.0013)^{NS}	0.0078	(0.0026)
DBP	0.0013	(0.0025)^{NS}	0.0009	(0.0013)^{NS}	0.0005	(0.0013)^{NS}	0.0118	(0.0025)
HR	0.0000	(0.0046)^{NS}	0.0028	(0.0024)^{NS}	0.0000	(0.0025)^{NS}	0.0230	(0.0047)

[a] Model '**KFC**' = $\mathbf{GRM_{kin}}$ + $\mathbf{ERM_{Family}}$ + $\mathbf{ERM_{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Couple}}$

[e] This column shows the residual variance

Trait	Model: KFC							
	V (s.e.)[f]		h^2_{kin} (s.e.)[g]		e^2_f (s.e.)[h]		e^2_c (s.e.)[i]	
Height	41.6630	(0.6680)	0.83	(0.17)	0.02	(0.08) NS	0.16	(0.09) NS
Weight	0.0340	(0.0005)	0.44	(0.17)	0.10	(0.09) NS	0.08	(0.09) NS
Fat	46.8875	(0.7212)	0.48	(0.17)	0.02	(0.09) NS	0.16	(0.09)
BMI	0.0296	(0.0005)	0.44	(0.17)	0.08	(0.09) NS	0.13	(0.09) NS
Hips	0.0087	(0.0001)	0.40	(0.18)	0.05	(0.09) NS	0.12	(0.09) NS
Waist	0.0183	(0.0003)	0.61	(0.17)	0.00	(0.09) NS	0.24	(0.09) NS
WHR	0.0064	(0.0001)	0.44	(0.17)	0.00	(0.09) NS	0.14	(0.09) NS
ABSI	0.0042	(0.0001)	0.37	(0.16)	0.00	(0.08) NS	0.09	(0.09) NS
Urea	0.0565	(0.0008)	0.12	(0.17) NS	0.09	(0.09) NS	0.04	(0.09) NS
Creatinine	0.0242	(0.0004)	0.65	(0.17)	0.03	(0.09) NS	0.13	(0.09) NS
Glucose	0.0131	(0.0002)	0.17	(0.17) NS	0.01	(0.09) NS	0.05	(0.09) NS
TC	0.0381	(0.0006)	0.29	(0.18)	0.05	(0.09) NS	0.03	(0.09) NS
HDL	0.0634	(0.0010)	0.65	(0.16)	0.00	(0.08) NS	0.18	(0.08) NS
SBP	0.0142	(0.0002)	0.33	(0.18) NS	0.00	(0.09) NS	0.12	(0.09) NS
DBP	0.0145	(0.0002)	0.09	(0.17) NS	0.06	(0.09) NS	0.03	(0.09) NS
HR	0.0258	(0.0004)	0.00	(0.18) NS	0.11	(0.09) NS	0.00	(0.10) NS

[f] This column shows the total phenotypic variance

[g] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM}_{kin}$

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{Family}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{Couple}$

Trait	Model: KFC				
		h^2_{gkin} (s.e.)[j]	%V_C[k]	logL[l]	n[m]
Height	0.83	(0.17)	100.00%	-20935.52	9,150
Weight	0.44	(0.17)	61.82%	10911.81	9,118
Fat	0.48	(0.17)	66.42%	-21500.79	8,926
BMI	0.44	(0.17)	64.48%	11515.82	9,107
Hips	0.40	(0.18)	55.74%	16791.95	8,984
Waist	0.61	(0.17)	85.12%	13524.11	9,016
WHR	0.44	(0.17)	58.47%	18145.58	8,995
ABSI	0.37	(0.16)	46.74%	19936.96	8,962
Urea	0.12	(0.17) [NS]	24.96%	8520.81	9,148
Creatinine	0.65	(0.17)	80.40%	12530.13	9,146
Glucose	0.17	(0.17) [NS]	22.18%	14808.62	8,936
TC	0.29	(0.18)	37.17%	10321.39	9,136
HDL	0.65	(0.16)	82.71%	8041.76	9,125
SBP	0.33	(0.18) [NS]	44.48%	14807.11	9,144
DBP	0.09	(0.17) [NS]	18.67%	14699.27	9,141
HR	0.00	(0.18) [NS]	10.86%	12053.91	9,126

[j] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[k] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[l] This column shows the log likelihood ratio for each analysis

[m] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (24) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'KSC'

Trait	Model: KSC[a]							
	σ^2_{kin} (s.e.)[b]		σ^2_{es} (s.e.)[c]		σ^2_{ec} (s.e.)[d]		σ^2_{ε} (s.e.)[e]	
Height	35.4385	(1.6589)	0.1026	(1.4803)[NS]	6.6869	(1.2803)	0.0001	(2.4365)
Weight	0.0213	(0.0014)	0.0003	(0.0013)[NS]	0.0061	(0.0010)	0.0064	(0.0019)
Fat	23.9960	(2.0488)	1.2216	(1.9209)[NS]	8.6199	(1.4243)	13.0498	(2.7740)
BMI	0.0173	(0.0013)	0.0000	(0.0011)[NS]	0.0061	(0.0009)	0.0061	(0.0017)
Hips	0.0042	(0.0004)	0.0000	(0.0003)[NS]	0.0014	(0.0003)	0.0031	(0.0005)
Waist	0.0095	(0.0008)	0.0000	(0.0007)[NS]	0.0036	(0.0005)	0.0052	(0.0011)
WHR	0.0022	(0.0003)	0.0000	(0.0003)[NS]	0.0005	(0.0002)	0.0037	(0.0004)
ABSI	0.0012	(0.0002)	0.0000	(0.0002)[NS]	0.0002	(0.0001)[NS]	0.0028	(0.0003)
Urea	0.0160	(0.0025)	0.0006	(0.0026)[NS]	0.0072	(0.0017)	0.0327	(0.0034)
Creatinine	0.0165	(0.0010)	0.0019	(0.0009)	0.0037	(0.0007)	0.0022	(0.0013)[NS]
Glucose	0.0021	(0.0006)	0.0012	(0.0007)[NS]	0.0007	(0.0004)	0.0090	(0.0008)
TC	0.0136	(0.0016)	0.0042	(0.0016)	0.0027	(0.0011)	0.0175	(0.0022)
HDL	0.0349	(0.0026)	0.0014	(0.0024)[NS]	0.0090	(0.0018)	0.0181	(0.0036)
SBP	0.0036	(0.0006)	0.0011	(0.0007)[NS]	0.0014	(0.0004)	0.0081	(0.0009)
DBP	0.0028	(0.0006)	0.0005	(0.0007)[NS]	0.0013	(0.0004)	0.0098	(0.0009)
HR	0.0083	(0.0011)	0.0005	(0.0012)[NS]	0.0024	(0.0008)	0.0147	(0.0016)

[a] Model '**KSC**' = $\mathbf{GRM}_{kin}$ + $\mathbf{ERM}_{Sib}$ + $\mathbf{ERM}_{Couple}$

[b] This column shows the variance captured by matrix $\mathbf{GRM}_{kin}$

[c] This column shows the variance captured by matrix $\mathbf{ERM}_{Sib}$

[d] This column shows the variance captured by matrix $\mathbf{ERM}_{Couple}$

[e] This column shows the residual variance

Trait	Model: KSC							
	V (s.e.)[f]		h^2_{kin} (s.e.)[g]		e^2_s (s.e.)[h]		e^2_c (s.e.)[i]	
Height	42.2281	(0.6802)	0.84	(0.03)	0.00	(0.04)[NS]	0.16	(0.03)
Weight	0.0340	(0.0005)	0.63	(0.04)	0.01	(0.04)[NS]	0.18	(0.03)
Fat	46.8874	(0.7211)	0.51	(0.04)	0.03	(0.04)[NS]	0.18	(0.03)
BMI	0.0296	(0.0005)	0.58	(0.04)	0.00	(0.04)[NS]	0.21	(0.03)
Hips	0.0087	(0.0001)	0.48	(0.04)	0.00	(0.04)[NS]	0.16	(0.03)
Waist	0.0183	(0.0003)	0.52	(0.04)	0.00	(0.04)[NS]	0.20	(0.03)
WHR	0.0064	(0.0001)	0.34	(0.04)	0.00	(0.04)[NS]	0.08	(0.03)
ABSI	0.0042	(0.0001)	0.28	(0.04)	0.00	(0.05)[NS]	0.04	(0.03)[NS]
Urea	0.0565	(0.0008)	0.28	(0.04)	0.01	(0.05)[NS]	0.13	(0.03)
Creatinine	0.0242	(0.0004)	0.68	(0.04)	0.08	(0.04)	0.15	(0.03)
Glucose	0.0131	(0.0002)	0.16	(0.04)	0.09	(0.05)[NS]	0.05	(0.03)
TC	0.0381	(0.0006)	0.36	(0.04)	0.11	(0.04)	0.07	(0.03)
HDL	0.0635	(0.0010)	0.55	(0.04)	0.02	(0.04)[NS]	0.14	(0.03)
SBP	0.0142	(0.0002)	0.25	(0.04)	0.08	(0.05)[NS]	0.10	(0.03)
DBP	0.0145	(0.0002)	0.19	(0.04)	0.04	(0.05)[NS]	0.09	(0.03)
HR	0.0259	(0.0004)	0.32	(0.04)	0.02	(0.05)[NS]	0.09	(0.03)

[f] This column shows the total phenotypic variance

[g] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM}_{kin}$

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{Sib}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{Couple}$

Trait	Model: KSC				
	h^2_{gkin} (s.e.)[j]		%V_C[k]	logL[l]	n[m]
Height	0.84	(0.03)	100.00%	-20945.68	9,150
Weight	0.63	(0.04)	81.28%	10911.13	9,118
Fat	0.51	(0.04)	72.17%	-21500.63	8,926
BMI	0.58	(0.04)	79.09%	11515.40	9,107
Hips	0.48	(0.04)	64.51%	16791.83	8,984
Waist	0.52	(0.04)	71.62%	13527.39	9,016
WHR	0.34	(0.04)	42.39%	18149.27	8,995
ABSI	0.28	(0.04)	33.12%	19939.60	8,962
Urea	0.28	(0.04)	42.19%	8520.25	9,148
Creatinine	0.68	(0.04)	91.06%	12532.22	9,146
Glucose	0.16	(0.04)	30.75%	14809.74	8,936
TC	0.36	(0.04)	53.98%	10324.69	9,136
HDL	0.55	(0.04)	71.40%	8045.64	9,125
SBP	0.25	(0.04)	43.08%	14809.20	9,144
DBP	0.19	(0.04)	31.81%	14699.35	9,141
HR	0.32	(0.04)	43.34%	12056.24	9,126

[j] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[k] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[l] This column shows the log likelihood ratio for each analysis

[m] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (25) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'FSC'

Trait	Model: FSC[a]							
	σ^2_{ef} (s.e.)[b]		σ^2_{es} (s.e.)[c]		σ^2_{ec} (s.e.)[d]		σ^2_{ε} (s.e.)[e]	
Height	14.3354	(0.8652)	0.0001	(1.4926)^{NS}	0.0001	(1.0838)^{NS}	23.9870	(1.6371)
Weight	0.0086	(0.0008)	0.0000	(0.0014)^{NS}	0.0000	(0.0011)^{NS}	0.0253	(0.0016)
Fat	10.9650	(1.0768)	0.1049	(2.0453)^{NS}	0.0001	(1.5347)^{NS}	35.7946	(2.3683)
BMI	0.0075	(0.0007)	0.0000	(0.0012)^{NS}	0.0000	(0.0009)^{NS}	0.0221	(0.0014)
Hips	0.0018	(0.0002)	0.0000	(0.0004)^{NS}	0.0000	(0.0003)^{NS}	0.0069	(0.0004)
Waist	0.0041	(0.0004)	0.0000	(0.0008)^{NS}	0.0000	(0.0006)^{NS}	0.0142	(0.0009)
WHR	0.0008	(0.0001)	0.0000	(0.0003)^{NS}	0.0000	(0.0002)^{NS}	0.0056	(0.0004)
ABSI	0.0003	(0.0001)	0.0000	(0.0002)^{NS}	0.0000	(0.0002)^{NS}	0.0038	(0.0002)
Urea	0.0080	(0.0013)	0.0003	(0.0027)^{NS}	0.0000	(0.0019)^{NS}	0.0485	(0.0031)
Creatinine	0.0068	(0.0006)	0.0012	(0.0012)^{NS}	0.0000	(0.0008)^{NS}	0.0176	(0.0013)
Glucose	0.0009	(0.0003)	0.0010	(0.0007)^{NS}	0.0000	(0.0005)^{NS}	0.0112	(0.0008)
TC	0.0056	(0.0009)	0.0029	(0.0017)	0.0000	(0.0013)^{NS}	0.0296	(0.0020)
HDL	0.0147	(0.0015)	0.0006	(0.0029)^{NS}	0.0000	(0.0022)^{NS}	0.0509	(0.0033)
SBP	0.0016	(0.0003)	0.0010	(0.0007)^{NS}	0.0000	(0.0005)^{NS}	0.0116	(0.0008)
DBP	0.0014	(0.0003)	0.0004	(0.0007)^{NS}	0.0000	(0.0005)^{NS}	0.0126	(0.0008)
HR	0.0038	(0.0006)	0.0000	(0.0013)^{NS}	0.0000	(0.0009)^{NS}	0.0227	(0.0015)

[a] Model '**FSC**' = $\mathbf{ERM}_{\mathbf{Family}}$ + $\mathbf{ERM}_{\mathbf{Sib}}$ + $\mathbf{ERM}_{\mathbf{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{ERM}_{\mathbf{Family}}$

[c] This column shows the variance captured by matrix $\mathbf{ERM}_{\mathbf{Sib}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM}_{\mathbf{Couple}}$

[e] This column shows the residual variance

Trait	Model: FSC					
	V (s.e.)[f]		e^2_f (s.e.)[g]		e^2_s (s.e.)[h]	
Height	38.3227	(0.5915)	0.37	(0.02)	0.00	(0.04)[NS]
Weight	0.0339	(0.0005)	0.25	(0.02)	0.00	(0.04)[NS]
Fat	46.8646	(0.7193)	0.23	(0.02)	0.00	(0.04)[NS]
BMI	0.0296	(0.0005)	0.25	(0.02)	0.00	(0.04)[NS]
Hips	0.0087	(0.0001)	0.20	(0.02)	0.00	(0.04)[NS]
Waist	0.0183	(0.0003)	0.22	(0.02)	0.00	(0.04)[NS]
WHR	0.0064	(0.0001)	0.12	(0.02)	0.00	(0.05)[NS]
ABSI	0.0042	(0.0001)	0.08	(0.02)	0.00	(0.05)[NS]
Urea	0.0568	(0.0009)	0.14	(0.02)	0.01	(0.05)[NS]
Creatinine	0.0257	(0.0004)	0.27	(0.02)	0.05	(0.05)[NS]
Glucose	0.0131	(0.0002)	0.07	(0.02)	0.08	(0.05)[NS]
TC	0.0380	(0.0006)	0.15	(0.02)	0.08	(0.05)
HDL	0.0662	(0.0010)	0.22	(0.02)	0.01	(0.04)[NS]
SBP	0.0142	(0.0002)	0.11	(0.02)	0.07	(0.05)[NS]
DBP	0.0145	(0.0002)	0.10	(0.02)	0.03	(0.05)[NS]
HR	0.0265	(0.0004)	0.14	(0.02)	0.00	(0.05)[NS]

[f] This column shows the total phenotypic variance

[g] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{\mathbf{Family}}$

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{\mathbf{Sib}}$

Trait	Model: FSC				
		e^2_c (s.e.)[i]	%V_C[j]	logL[k]	n[l]
Height	0.00	(0.03) [NS]	37.41%	-20975.03	9,150
Weight	0.00	(0.03) [NS]	25.27%	10898.30	9,118
Fat	0.00	(0.03) [NS]	23.62%	-21506.90	8,926
BMI	0.00	(0.03) [NS]	25.29%	11508.69	9,107
Hips	0.00	(0.03) [NS]	20.17%	16786.49	8,984
Waist	0.00	(0.03) [NS]	22.38%	13519.33	9,016
WHR	0.00	(0.04) [NS]	11.99%	18142.30	8,995
ABSI	0.00	(0.04) [NS]	8.23%	19932.37	8,962
Urea	0.00	(0.03) [NS]	14.63%	8520.33	9,148
Creatinine	0.00	(0.03) [NS]	31.25%	12493.88	9,146
Glucose	0.00	(0.04) [NS]	14.64%	14808.70	8,936
TC	0.00	(0.03) [NS]	22.23%	10317.28	9,136
HDL	0.00	(0.03) [NS]	23.08%	8023.67	9,125
SBP	0.00	(0.03) [NS]	18.47%	14807.33	9,144
DBP	0.00	(0.03) [NS]	12.47%	14699.31	9,141
HR	0.00	(0.03) [NS]	14.19%	12054.52	9,126

[i] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Couple}**

[j] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[k] This column shows the log likelihood ratio for each analysis

[l] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (26) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GKFS'

Trait	Model: GKFS[a]							
	σ^2_g (s.e.)[b]		σ^2_{kin} (s.e.)[c]		σ^2_{ef} (s.e.)[d]		σ^2_{es} (s.e.)[e]	
Height	20.8251	(1.5662)	0.1066	(2.1469)^{NS}	8.0840	(0.9921)	0.6769	(1.0363)^{NS}
Weight	0.0095	(0.0014)	0.0004	(0.0023)^{NS}	0.0060	(0.0010)	0.0002	(0.0012)^{NS}
Fat	11.5330	(1.9583)	0.0001	(3.3301)^{NS}	7.0046	(1.3946)	0.0362	(1.8818)^{NS}
BMI	0.0070	(0.0012)	0.0000	(0.0020)^{NS}	0.0054	(0.0008)	0.0000	(0.0011)^{NS}
Hips	0.0017	(0.0004)	0.0000	(0.0006)^{NS}	0.0013	(0.0003)	0.0000	(0.0003)^{NS}
Waist	0.0029	(0.0007)	0.0000	(0.0013)^{NS}	0.0032	(0.0005)	0.0000	(0.0007)^{NS}
WHR	0.0009	(0.0003)	0.0003	(0.0005)^{NS}	0.0004	(0.0002)	0.0000	(0.0003)^{NS}
ABSI	0.0004	(0.0002)	0.0005	(0.0003)^{NS}	0.0001	(0.0001)^{NS}	0.0000	(0.0002)^{NS}
Urea	0.0074	(0.0023)	0.0000	(0.0043)^{NS}	0.0056	(0.0017)	0.0001	(0.0027)^{NS}
Creatinine	0.0059	(0.0010)	0.0035	(0.0016)	0.0037	(0.0007)	0.0016	(0.0009)
Glucose	0.0022	(0.0005)	0.0000	(0.0010)^{NS}	0.0002	(0.0004)^{NS}	0.0006	(0.0007)^{NS}
TC	0.0059	(0.0015)	0.0027	(0.0027)^{NS}	0.0027	(0.0011)	0.0044	(0.0016)
HDL	0.0190	(0.0026)	0.0003	(0.0043)^{NS}	0.0080	(0.0018)	0.0008	(0.0024)^{NS}
SBP	0.0019	(0.0006)	0.0000	(0.0010)^{NS}	0.0011	(0.0004)	0.0009	(0.0007)^{NS}
DBP	0.0016	(0.0006)	0.0000	(0.0011)^{NS}	0.0009	(0.0004)	0.0001	(0.0007)^{NS}
HR	0.0037	(0.0010)	0.0000	(0.0019)^{NS}	0.0026	(0.0007)	0.0001	(0.0012)^{NS}

[a] Model '**GKFS**' = $\mathbf{GRM_g}$ +$\mathbf{GRM_{kin}}$ + $\mathbf{ERM_{Family}}$ + $\mathbf{ERM_{Sib}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[e] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

Trait	Model: GKFS									
	σ^2_ε (s.e.)[f]		V (s.e.)[g]		h^2_g (s.e.)[h]		h^2_{kin} (s.e.)[i]		e^2_f (s.e.)[j]	
Height	8.9879	(1.4329)	38.6806	(0.6149)	0.54	(0.04)	0.00	(0.06) [NS]	0.21	(0.03)
Weight	0.0180	(0.0017)	0.0341	(0.0005)	0.28	(0.04)	0.01	(0.07) [NS]	0.18	(0.03)
Fat	28.3743	(2.4539)	46.9483	(0.7255)	0.25	(0.04)	0.00	(0.07) [NS]	0.15	(0.03)
BMI	0.0172	(0.0015)	0.0296	(0.0005)	0.24	(0.04)	0.00	(0.07) [NS]	0.18	(0.03)
Hips	0.0057	(0.0005)	0.0087	(0.0001)	0.20	(0.04)	0.00	(0.07) [NS]	0.15	(0.03)
Waist	0.0121	(0.0010)	0.0183	(0.0003)	0.16	(0.04)	0.00	(0.07) [NS]	0.18	(0.03)
WHR	0.0047	(0.0004)	0.0064	(0.0001)	0.14	(0.04)	0.05	(0.07) [NS]	0.07	(0.03)
ABSI	0.0032	(0.0002)	0.0042	(0.0001)	0.09	(0.04)	0.11	(0.08) [NS]	0.03	(0.03) [NS]
Urea	0.0450	(0.0033)	0.0581	(0.0009)	0.13	(0.04)	0.00	(0.07) [NS]	0.10	(0.03)
Creatinine	0.0095	(0.0011)	0.0242	(0.0004)	0.24	(0.04)	0.14	(0.06)	0.15	(0.03)
Glucose	0.0100	(0.0008)	0.0131	(0.0002)	0.17	(0.04)	0.00	(0.07) [NS]	0.02	(0.03) [NS]
TC	0.0225	(0.0019)	0.0381	(0.0006)	0.15	(0.04)	0.07	(0.07) [NS]	0.07	(0.03)
HDL	0.0354	(0.0032)	0.0636	(0.0010)	0.30	(0.04)	0.00	(0.07) [NS]	0.13	(0.03)
SBP	0.0104	(0.0008)	0.0142	(0.0002)	0.13	(0.04)	0.00	(0.07) [NS]	0.08	(0.03)
DBP	0.0119	(0.0008)	0.0145	(0.0002)	0.11	(0.04)	0.00	(0.07) [NS]	0.06	(0.03)
HR	0.0196	(0.0015)	0.0259	(0.0004)	0.14	(0.04)	0.00	(0.07) [NS]	0.10	(0.03)

[f] This column shows the residual variance

[g] This column shows the total phenotypic variance

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_g}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_{kin}}$

[j] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Family}}$

Trait	Model: GKFS						
	e^2_s (s.e.)[k]		h^2_{gkin} (s.e.)[l]		%V_C[m]	logL[n]	n[o]
Height	0.02	(0.03) [NS]	0.54	(0.05)	76.76%	-20805.17	9,150
Weight	0.01	(0.04) [NS]	0.29	(0.06)	47.10%	10936.83	9,118
Fat	0.00	(0.04) [NS]	0.25	(0.06)	39.56%	-21481.57	8,926
BMI	0.00	(0.04) [NS]	0.24	(0.06)	41.78%	11535.98	9,107
Hips	0.00	(0.04) [NS]	0.20	(0.06)	34.48%	16805.32	8,984
Waist	0.00	(0.04) [NS]	0.16	(0.06)	33.69%	13531.87	9,016
WHR	0.00	(0.04) [NS]	0.19	(0.06)	25.58%	18155.07	8,995
ABSI	0.00	(0.05) [NS]	0.20	(0.06)	23.03%	19941.83	8,962
Urea	0.00	(0.05) [NS]	0.13	(0.06)	22.51%	8526.41	9,148
Creatinine	0.07	(0.04)	0.39	(0.05)	60.75%	12550.94	9,146
Glucose	0.05	(0.05) [NS]	0.17	(0.06)	22.95%	14820.77	8,936
TC	0.12	(0.04)	0.22	(0.06)	41.01%	10332.95	9,136
HDL	0.01	(0.04) [NS]	0.30	(0.06)	44.27%	8072.58	9,125
SBP	0.06	(0.05) [NS]	0.13	(0.06)	26.94%	14815.04	9,144
DBP	0.01	(0.05) [NS]	0.11	(0.06)	17.95%	14704.68	9,141
HR	0.00	(0.05) [NS]	0.14	(0.06)	24.43%	12064.95	9,126

[k] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Sib}**

[l] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[m] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[n] This column shows the log likelihood ratio for each analysis

[o] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (27) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GKFC'

Trait	Model: GKFC[a]							
	σ^2_g (s.e.)[b]		σ^2_{kin} (s.e.)[c]		σ^2_{ef} (s.e.)[d]		σ^2_{ec} (s.e.)[e]	
Height	17.9933	(1.5772)	14.2564	(6.2440)^{NS}	0.0001	(3.0749)^{NS}	6.3495	(3.1649)^{NS}
Weight	0.0095	(0.0014)	0.0056	(0.0059)^{NS}	0.0033	(0.0029)^{NS}	0.0029	(0.0030)^{NS}
Fat	12.3799	(1.9589)	9.8104	(8.2481)^{NS}	1.2045	(4.0793)^{NS}	7.6018	(4.2216)
BMI	0.0075	(0.0012)	0.0058	(0.0051)^{NS}	0.0021	(0.0026)^{NS}	0.0041	(0.0026)
Hips	0.0018	(0.0004)	0.0016	(0.0016)^{NS}	0.0004	(0.0008)^{NS}	0.0010	(0.0008)^{NS}
Waist	0.0027	(0.0007)	0.0080	(0.0032)^{NS}	0.0000	(0.0016)^{NS}	0.0044	(0.0016)
WHR	0.0009	(0.0003)	0.0019	(0.0011)^{NS}	0.0000	(0.0006)^{NS}	0.0009	(0.0006)^{NS}
ABSI	0.0004	(0.0002)	0.0012	(0.0007)^{NS}	0.0000	(0.0004)^{NS}	0.0004	(0.0004)^{NS}
Urea	0.0084	(0.0022)	0.0000	(0.0097)^{NS}	0.0044	(0.0049)^{NS}	0.0019	(0.0051)^{NS}
Creatinine	0.0059	(0.0010)	0.0087	(0.0043)	0.0013	(0.0021)^{NS}	0.0028	(0.0022)^{NS}
Glucose	0.0025	(0.0005)	0.0000	(0.0022)^{NS}	0.0000	(0.0011)^{NS}	0.0006	(0.0012)^{NS}
TC	0.0057	(0.0015)	0.0049	(0.0068)	0.0023	(0.0034)^{NS}	0.0008	(0.0035)^{NS}
HDL	0.0182	(0.0026)	0.0220	(0.0105)^{NS}	0.0000	(0.0052)^{NS}	0.0119	(0.0053)^{NS}
SBP	0.0020	(0.0006)	0.0026	(0.0026)^{NS}	0.0000	(0.0013)^{NS}	0.0018	(0.0013)^{NS}
DBP	0.0018	(0.0006)	0.0000	(0.0025)^{NS}	0.0007	(0.0013)^{NS}	0.0003	(0.0013)^{NS}
HR	0.0008	(0.0010)	0.0000	(0.0047)^{NS}	0.0025	(0.0024)^{NS}	0.0000	(0.0025)^{NS}

[a] Model '**GKFC**' = $\mathbf{GRM_g}$ +$\mathbf{GRM_{kin}}$ + $\mathbf{ERM_{Family}}$ + $\mathbf{ERM_{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[e] This column shows the variance captured by matrix $\mathbf{ERM_{Couple}}$

Trait	Model: GKFC									
	σ^2_ε (s.e.)[f]		V (s.e.)[g]		h^2_g (s.e.)[h]		h^2_{kin} (s.e.)[i]		e^2_f (s.e.)[j]	
Height	0.0001	(6.1458) [NS]	38.5995	(0.6019)	0.47	(0.04)	0.37	(0.16) [NS]	0.00	(0.08) [NS]
Weight	0.0127	(0.0058)	0.0341	(0.0005)	0.28	(0.04)	0.16	(0.17) [NS]	0.10	(0.09) [NS]
Fat	15.9676	(8.1172)	46.9641	(0.7269)	0.26	(0.04)	0.21	(0.18) [NS]	0.03	(0.09) [NS]
BMI	0.0101	(0.0051)	0.0296	(0.0005)	0.25	(0.04)	0.19	(0.17) [NS]	0.07	(0.09) [NS]
Hips	0.0038	(0.0016)	0.0087	(0.0001)	0.21	(0.04)	0.18	(0.18) [NS]	0.05	(0.09) [NS]
Waist	0.0031	(0.0031) [NS]	0.0183	(0.0003)	0.15	(0.04)	0.44	(0.17) [NS]	0.00	(0.09) [NS]
WHR	0.0028	(0.0011)	0.0064	(0.0001)	0.14	(0.04)	0.29	(0.17) [NS]	0.00	(0.09) [NS]
ABSI	0.0023	(0.0007)	0.0042	(0.0001)	0.08	(0.04)	0.27	(0.17) [NS]	0.00	(0.08) [NS]
Urea	0.0417	(0.0097)	0.0565	(0.0008)	0.15	(0.04)	0.00	(0.17) [NS]	0.08	(0.09) [NS]
Creatinine	0.0056	(0.0042) [NS]	0.0242	(0.0004)	0.24	(0.04)	0.36	(0.18)	0.05	(0.09) [NS]
Glucose	0.0100	(0.0022)	0.0132	(0.0002)	0.19	(0.04)	0.00	(0.17) [NS]	0.00	(0.08) [NS]
TC	0.0245	(0.0067)	0.0381	(0.0006)	0.15	(0.04)	0.13	(0.18)	0.06	(0.09) [NS]
HDL	0.0115	(0.0103) [NS]	0.0635	(0.0010)	0.29	(0.04)	0.35	(0.17) [NS]	0.00	(0.08) [NS]
SBP	0.0079	(0.0025)	0.0142	(0.0002)	0.14	(0.04)	0.18	(0.18) [NS]	0.00	(0.09) [NS]
DBP	0.0116	(0.0025)	0.0145	(0.0002)	0.13	(0.04)	0.00	(0.18) [NS]	0.05	(0.09) [NS]
HR	0.0225	(0.0047)	0.0258	(0.0004)	0.03	(0.04)	0.00	(0.18) [NS]	0.10	(0.09) [NS]

[f] This column shows the residual variance

[g] This column shows the total phenotypic variance

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_g}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_{kin}}$

[j] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Family}}$

Trait	Model: GKFC						
	e^2_c (s.e.)[k]		h^2_{gkin} (s.e.)[l]		%V_C[m]	logL[n]	n[o]
Height	0.16	(0.08) [NS]	0.84	(0.12)	100.00%	-20814.98	9,150
Weight	0.09	(0.09) [NS]	0.44	(0.12)	62.44%	10937.31	9,118
Fat	0.16	(0.09)	0.47	(0.13)	66.00%	-21479.36	8,926
BMI	0.14	(0.09)	0.44	(0.12)	65.95%	11537.52	9,107
Hips	0.12	(0.09) [NS]	0.39	(0.13)	55.63%	16806.27	8,984
Waist	0.24	(0.09)	0.59	(0.12)	82.82%	13533.64	9,016
WHR	0.14	(0.09) [NS]	0.43	(0.12)	57.39%	18153.61	8,995
ABSI	0.09	(0.09) [NS]	0.35	(0.12)	46.54%	19940.55	8,962
Urea	0.03	(0.09) [NS]	0.15	(0.12)	26.06%	8529.18	9,148
Creatinine	0.12	(0.09) [NS]	0.60	(0.13)	77.17%	12550.10	9,146
Glucose	0.04	(0.09) [NS]	0.19	(0.12)	23.74%	14821.15	8,936
TC	0.02	(0.09) [NS]	0.28	(0.13)	35.78%	10329.11	9,136
HDL	0.19	(0.08) [NS]	0.64	(0.12)	82.07%	8072.28	9,125
SBP	0.12	(0.09) [NS]	0.32	(0.13)	44.94%	14814.47	9,144
DBP	0.02	(0.09) [NS]	0.13	(0.13)	19.64%	14705.25	9,141
HR	0.00	(0.10) [NS]	0.03	(0.13)	12.79%	12057.17	9,126

[k] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Couple}**

[l] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[m] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[n] This column shows the log likelihood ratio for each analysis

[o] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (28) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GKSC'

Trait	Model: GKSC[a]							
	σ^2_g (s.e.)[b]		σ^2_{kin} (s.e.)[c]		σ^2_{es} (s.e.)[d]		σ^2_{ec} (s.e.)[e]	
Height	19.0108	(1.7280)	14.5253	(2.2142)	0.6836	(1.4160)^{NS}	6.8039	(1.1936)
Weight	0.0095	(0.0014)	0.0119	(0.0019)	0.0003	(0.0012)^{NS}	0.0061	(0.0010)
Fat	12.3689	(1.9588)	11.7899	(2.7107)	1.1370	(1.9036)^{NS}	8.6849	(1.4098)
BMI	0.0075	(0.0012)	0.0098	(0.0017)	0.0000	(0.0011)^{NS}	0.0062	(0.0009)
Hips	0.0018	(0.0004)	0.0024	(0.0005)	0.0000	(0.0003)^{NS}	0.0015	(0.0003)
Waist	0.0029	(0.0007)	0.0065	(0.0011)	0.0000	(0.0007)^{NS}	0.0037	(0.0005)
WHR	0.0009	(0.0003)	0.0012	(0.0004)	0.0000	(0.0003)^{NS}	0.0005	(0.0002)
ABSI	0.0004	(0.0002)	0.0008	(0.0002)	0.0000	(0.0002)^{NS}	0.0002	(0.0001)
Urea	0.0090	(0.0022)	0.0070	(0.0032)	0.0005	(0.0026)^{NS}	0.0073	(0.0017)
Creatinine	0.0059	(0.0010)	0.0108	(0.0013)	0.0018	(0.0008)	0.0038	(0.0007)
Glucose	0.0025	(0.0005)	0.0000	(0.0008)^{NS}	0.0011	(0.0007)^{NS}	0.0006	(0.0004)
TC	0.0058	(0.0015)	0.0079	(0.0022)	0.0045	(0.0016)	0.0027	(0.0011)
HDL	0.0190	(0.0026)	0.0161	(0.0035)	0.0011	(0.0024)^{NS}	0.0093	(0.0018)
SBP	0.0021	(0.0006)	0.0015	(0.0008)	0.0012	(0.0007)^{NS}	0.0014	(0.0004)
DBP	0.0020	(0.0006)	0.0008	(0.0008)^{NS}	0.0005	(0.0007)^{NS}	0.0013	(0.0004)
HR	0.0038	(0.0010)	0.0046	(0.0015)	0.0004	(0.0012)^{NS}	0.0024	(0.0008)

[a] Model '**GKSC**' = $\mathbf{GRM_g}$ +$\mathbf{GRM_{kin}}$ + $\mathbf{ERM_{Sib}}$ + $\mathbf{ERM_{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

[e] This column shows the variance captured by matrix $\mathbf{ERM_{Couple}}$

Trait	Model: GKSC									
	σ^2_ε (s.e.)[f]		V (s.e.)[g]		h^2_g (s.e.)[h]		h^2_{kin} (s.e.)[i]		e^2_s (s.e.)[j]	
Height	0.0001	(2.2663) [NS]	41.0237	(0.6587)	0.46	(0.04)	0.35	(0.05)	0.02	(0.03) [NS]
Weight	0.0063	(0.0019)	0.0340	(0.0005)	0.28	(0.04)	0.35	(0.05)	0.01	(0.04) [NS]
Fat	12.9818	(2.7432)	46.9624	(0.7268)	0.26	(0.04)	0.25	(0.06)	0.02	(0.04) [NS]
BMI	0.0060	(0.0017)	0.0296	(0.0005)	0.25	(0.04)	0.33	(0.06)	0.00	(0.04) [NS]
Hips	0.0030	(0.0005)	0.0087	(0.0001)	0.21	(0.04)	0.27	(0.06)	0.00	(0.04) [NS]
Waist	0.0052	(0.0011)	0.0183	(0.0003)	0.16	(0.04)	0.36	(0.06)	0.00	(0.04) [NS]
WHR	0.0037	(0.0004)	0.0064	(0.0001)	0.14	(0.04)	0.19	(0.06)	0.00	(0.04) [NS]
ABSI	0.0028	(0.0003)	0.0042	(0.0001)	0.09	(0.04)	0.19	(0.06)	0.00	(0.05) [NS]
Urea	0.0327	(0.0034)	0.0565	(0.0008)	0.16	(0.04)	0.12	(0.06)	0.01	(0.05) [NS]
Creatinine	0.0020	(0.0013) [NS]	0.0242	(0.0004)	0.24	(0.04)	0.45	(0.05)	0.07	(0.03)
Glucose	0.0089	(0.0008)	0.0131	(0.0002)	0.19	(0.04)	0.00	(0.06) [NS]	0.08	(0.05) [NS]
TC	0.0172	(0.0022)	0.0381	(0.0006)	0.15	(0.04)	0.21	(0.06)	0.12	(0.04)
HDL	0.0181	(0.0036)	0.0636	(0.0010)	0.30	(0.04)	0.25	(0.05)	0.02	(0.04) [NS]
SBP	0.0080	(0.0008)	0.0142	(0.0002)	0.15	(0.04)	0.11	(0.06)	0.08	(0.05) [NS]
DBP	0.0098	(0.0009)	0.0145	(0.0002)	0.14	(0.04)	0.06	(0.06) [NS]	0.04	(0.05) [NS]
HR	0.0147	(0.0016)	0.0259	(0.0004)	0.15	(0.04)	0.18	(0.06)	0.02	(0.05) [NS]

[f] This column shows the residual variance

[g] This column shows the total phenotypic variance

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_g}$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM_{kin}}$

[j] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Sib}}$

Trait	Model: GKSC						
	e^2_c (s.e.)[k]		h^2_{gkin} (s.e.)[l]		%V_C[m]	logL[n]	n[o]
Height	0.17	(0.03)	0.82	(0.05)	100.00%	-20831.87	9,150
Weight	0.18	(0.03)	0.63	(0.05)	81.57%	10936.65	9,118
Fat	0.18	(0.03)	0.51	(0.05)	72.36%	-21479.23	8,926
BMI	0.21	(0.03)	0.59	(0.05)	79.51%	11537.17	9,107
Hips	0.17	(0.03)	0.48	(0.05)	65.19%	16806.12	8,984
Waist	0.20	(0.03)	0.51	(0.05)	71.72%	13535.79	9,016
WHR	0.08	(0.03)	0.34	(0.05)	42.17%	18156.38	8,995
ABSI	0.05	(0.03)	0.28	(0.05)	32.70%	19942.56	8,962
Urea	0.13	(0.03)	0.28	(0.05)	42.12%	8528.96	9,148
Creatinine	0.16	(0.03)	0.69	(0.05)	91.84%	12551.88	9,146
Glucose	0.04	(0.03)	0.19	(0.05)	31.91%	14822.19	8,936
TC	0.07	(0.03)	0.36	(0.05)	54.86%	10332.84	9,136
HDL	0.15	(0.03)	0.55	(0.05)	71.63%	8075.23	9,125
SBP	0.10	(0.03)	0.25	(0.05)	43.94%	14816.60	9,144
DBP	0.09	(0.03)	0.20	(0.05)	31.82%	14705.77	9,141
HR	0.09	(0.03)	0.32	(0.05)	43.20%	12063.28	9,126

[k] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Couple}**

[l] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[m] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[n] This column shows the log likelihood ratio for each analysis

[o] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (29) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GFSC'

Trait	Model: GFSC[a]							
	σ^2_g (s.e.)[b]		σ^2_{ef} (s.e.)[c]		σ^2_{es} (s.e.)[d]		σ^2_{ec} (s.e.)[e]	
Height	21.8140	(1.5351)	7.6165	(0.9293)	0.5938	(1.0461)^{NS}	1.2357	(1.0705)^{NS}
Weight	0.0098	(0.0014)	0.0060	(0.0009)	0.0002	(0.0012)^{NS}	0.0003	(0.0012)^{NS}
Fat	12.9125	(1.9159)	5.6387	(1.3448)	1.0547	(1.9187)^{NS}	3.1171	(1.7245)
BMI	0.0078	(0.0012)	0.0049	(0.0008)	0.0000	(0.0011)^{NS}	0.0014	(0.0010)^{NS}
Hips	0.0019	(0.0003)	0.0012	(0.0002)	0.0000	(0.0003)^{NS}	0.0003	(0.0003)^{NS}
Waist	0.0034	(0.0007)	0.0029	(0.0005)	0.0000	(0.0008)^{NS}	0.0007	(0.0007)^{NS}
WHR	0.0010	(0.0003)	0.0005	(0.0002)	0.0000	(0.0003)^{NS}	0.0000	(0.0003)^{NS}
ABSI	0.0004	(0.0002)	0.0003	(0.0001)	0.0000	(0.0002)^{NS}	0.0000	(0.0002)^{NS}
Urea	0.0088	(0.0022)	0.0039	(0.0016)	0.0003	(0.0026)^{NS}	0.0035	(0.0022)^{NS}
Creatinine	0.0060	(0.0009)	0.0048	(0.0007)	0.0012	(0.0009)	0.0000	(0.0008)^{NS}
Glucose	0.0025	(0.0005)	0.0000	(0.0004)^{NS}	0.0012	(0.0007)^{NS}	0.0008	(0.0005)^{NS}
TC	0.0058	(0.0015)	0.0036	(0.0011)	0.0042	(0.0016)	0.0000	(0.0014)^{NS}
HDL	0.0204	(0.0025)	0.0072	(0.0017)	0.0009	(0.0025)^{NS}	0.0021	(0.0022)^{NS}
SBP	0.0022	(0.0006)	0.0007	(0.0004)^{NS}	0.0012	(0.0007)	0.0008	(0.0005)^{NS}
DBP	0.0020	(0.0006)	0.0005	(0.0004)^{NS}	0.0005	(0.0007)^{NS}	0.0008	(0.0006)^{NS}
HR	0.0035	(0.0010)	0.0026	(0.0007)	0.0001	(0.0012)^{NS}	0.0000	(0.0010)^{NS}

[a] Model '**GFSC**' = $\mathbf{GRM_g}$ +$\mathbf{ERM_{Family}}$ + $\mathbf{ERM_{Sib}}$ + $\mathbf{ERM_{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

[e] This column shows the variance captured by matrix $\mathbf{ERM_{Couple}}$

Trait	Model: GFSC									
	σ^2_ε (s.e.)[f]		V (s.e.)[g]		h^2_g (s.e.)[h]		e^2_f (s.e.)[i]		e^2_s (s.e.)[j]	
Height	7.4829	(1.8147)	38.7429	(0.6178)	0.56	(0.04)	0.20	(0.02)	0.02	(0.03) [NS]
Weight	0.0179	(0.0019)	0.0341	(0.0005)	0.29	(0.04)	0.18	(0.03)	0.00	(0.04) [NS]
Fat	24.2511	(2.8646)	46.9740	(0.7270)	0.27	(0.04)	0.12	(0.03)	0.02	(0.04) [NS]
BMI	0.0156	(0.0017)	0.0296	(0.0005)	0.26	(0.04)	0.16	(0.03)	0.00	(0.04) [NS]
Hips	0.0054	(0.0005)	0.0087	(0.0001)	0.22	(0.04)	0.13	(0.03)	0.00	(0.04) [NS]
Waist	0.0113	(0.0011)	0.0183	(0.0003)	0.18	(0.04)	0.16	(0.03)	0.00	(0.04) [NS]
WHR	0.0049	(0.0004)	0.0064	(0.0001)	0.16	(0.04)	0.08	(0.03)	0.00	(0.05) [NS]
ABSI	0.0035	(0.0003)	0.0042	(0.0001)	0.09	(0.04)	0.07	(0.03)	0.00	(0.05) [NS]
Urea	0.0400	(0.0037)	0.0565	(0.0008)	0.16	(0.04)	0.07	(0.03)	0.00	(0.05) [NS]
Creatinine	0.0121	(0.0014)	0.0242	(0.0004)	0.25	(0.04)	0.20	(0.03)	0.05	(0.04)
Glucose	0.0086	(0.0009)	0.0131	(0.0002)	0.19	(0.04)	0.00	(0.03) [NS]	0.09	(0.05) [NS]
TC	0.0245	(0.0024)	0.0381	(0.0006)	0.15	(0.04)	0.10	(0.03)	0.11	(0.04)
HDL	0.0331	(0.0037)	0.0636	(0.0010)	0.32	(0.04)	0.11	(0.03)	0.01	(0.04) [NS]
SBP	0.0093	(0.0009)	0.0142	(0.0002)	0.16	(0.04)	0.05	(0.03) [NS]	0.09	(0.05)
DBP	0.0107	(0.0009)	0.0145	(0.0002)	0.14	(0.04)	0.03	(0.03) [NS]	0.03	(0.05) [NS]
HR	0.0197	(0.0017)	0.0259	(0.0004)	0.13	(0.04)	0.10	(0.03)	0.00	(0.05) [NS]

[f] This column shows the residual variance

[g] This column shows the total phenotypic variance

[h] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{GRM}_g$

[i] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{Family}$

[j] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM}_{Sib}$

Trait	Model: GFSC						
	e^2_c (s.e.)[k]		h^2_{gkin} (s.e.)[l]		%V_C[m]	logL[n]	n[o]
Height	0.03	(0.03) [NS]	0.56	(0.04)	80.69%	-20804.51	9,150
Weight	0.01	(0.03) [NS]	0.29	(0.04)	47.64%	10936.84	9,118
Fat	0.07	(0.04)	0.27	(0.04)	48.37%	-21479.94	8,926
BMI	0.05	(0.03) [NS]	0.26	(0.04)	47.46%	11536.89	9,107
Hips	0.03	(0.04) [NS]	0.22	(0.04)	38.67%	16805.79	8,984
Waist	0.04	(0.04) [NS]	0.18	(0.04)	38.23%	13532.56	9,016
WHR	0.00	(0.04) [NS]	0.16	(0.04)	24.20%	18154.81	8,995
ABSI	0.00	(0.04) [NS]	0.09	(0.04)	16.05%	19940.06	8,962
Urea	0.06	(0.04) [NS]	0.16	(0.04)	29.30%	8529.46	9,148
Creatinine	0.00	(0.03) [NS]	0.25	(0.04)	49.85%	12547.07	9,146
Glucose	0.06	(0.04) [NS]	0.19	(0.04)	34.62%	14822.15	8,936
TC	0.00	(0.04) [NS]	0.15	(0.04)	35.64%	10332.17	9,136
HDL	0.03	(0.03) [NS]	0.32	(0.04)	47.97%	8073.01	9,125
SBP	0.05	(0.04) [NS]	0.16	(0.04)	34.98%	14816.19	9,144
DBP	0.06	(0.04) [NS]	0.14	(0.04)	26.09%	14705.90	9,141
HR	0.00	(0.04) [NS]	0.13	(0.04)	23.96%	12064.94	9,126

[k] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Couple}**

[l] This column shows the heritability estimate, which is the sum of h^2_g and h^2_{kin}

[m] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[n] This column shows the log likelihood ratio for each analysis

[o] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (30) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'KFSC'

Trait	Model: KFSC[a]							
	σ^2_{kin} (s.e.)[b]		σ^2_{ef} (s.e.)[c]		σ^2_{es} (s.e.)[d]		σ^2_{ec} (s.e.)[e]	
Height	35.4148	(7.4007)	0.7448	(3.7339)[NS]	0.0370	(1.5337)[NS]	7.1764	(3.8305)[NS]
Weight	0.0148	(0.0058)	0.0034	(0.0029)[NS]	0.0002	(0.0013)[NS]	0.0028	(0.003)[NS]
Fat	22.4800	(8.1083)	0.7990	(4.1214)[NS]	1.1863	(1.9282)[NS]	7.8462	(4.2597)
BMI	0.0129	(0.0051)	0.0023	(0.0026)[NS]	0.0000	(0.0011)[NS]	0.0039	(0.0027)[NS]
Hips	0.0034	(0.0016)	0.0004	(0.0008)[NS]	0.0000	(0.0003)[NS]	0.0010	(0.0008)[NS]
Waist	0.0111	(0.0031)	0.0000	(0.0016)[NS]	0.0000	(0.0007)[NS]	0.0043	(0.0016)[NS]
WHR	0.0028	(0.0011)	0.0000	(0.0006)[NS]	0.0000	(0.0003)[NS]	0.0008	(0.0006)[NS]
ABSI	0.0015	(0.0007)	0.0000	(0.0004)[NS]	0.0000	(0.0002)[NS]	0.0003	(0.0004)[NS]
Urea	0.0067	(0.0095)[NS]	0.0050	(0.0049)[NS]	0.0003	(0.0026)[NS]	0.0023	(0.0051)[NS]
Creatinine	0.0164	(0.0042)	0.0001	(0.0021)[NS]	0.0018	(0.0009)	0.0036	(0.0021)[NS]
Glucose	0.0022	(0.0022)[NS]	0.0000	(0.0011)[NS]	0.0012	(0.0007)[NS]	0.0007	(0.0012)[NS]
TC	0.0108	(0.0067)	0.0015	(0.0034)[NS]	0.0042	(0.0016)	0.0013	(0.0035)[NS]
HDL	0.0416	(0.0102)	0.0000	(0.0052)[NS]	0.0005	(0.0022)[NS]	0.0118	(0.0053)[NS]
SBP	0.0046	(0.0025)[NS]	0.0000	(0.0013)[NS]	0.0010	(0.0006)[NS]	0.0018	(0.0013)[NS]
DBP	0.0014	(0.0025)[NS]	0.0007	(0.0013)[NS]	0.0005	(0.0007)[NS]	0.0006	(0.0013)[NS]
HR	-[p]	-	-	-	-	-	-	-

[a] Model '**KFSC**' = $\mathbf{GRM_{kin}}$ +$\mathbf{ERM_{Family}}$ + $\mathbf{ERM_{Sib}}$ + $\mathbf{ERM_{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[c] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

[e] This column shows the variance captured by matrix $\mathbf{ERM_{Couple}}$

Trait	Model: KFSC									
	σ^2_ε (s.e.)[f]		V (s.e.)[g]		h^2_{kin} (s.e.)[h]		e^2_f (s.e.)[i]		e^2_s (s.e.)[j]	
Height	0.0001	(7.6002) NS	43.3731	(0.7092)	0.82	(0.17)	0.02	(0.09) NS	0.00	(0.04) NS
Weight	0.0128	(0.0059)	0.0340	(0.0005)	0.44	(0.17)	0.10	(0.09) NS	0.01	(0.04) NS
Fat	14.5774	(8.3938) NS	46.8890	(0.7212)	0.48	(0.17)	0.02	(0.09) NS	0.03	(0.04) NS
BMI	0.0105	(0.0052)	0.0296	(0.0005)	0.43	(0.17)	0.08	(0.09) NS	0.00	(0.04) NS
Hips	0.0039	(0.0016)	0.0087	(0.0001)	0.39	(0.18)	0.05	(0.09) NS	0.00	(0.04) NS
Waist	0.0029	(0.0032) NS	0.0183	(0.0003)	0.61	(0.17)	0.00	(0.09) NS	0.00	(0.04) NS
WHR	0.0028	(0.0011)	0.0064	(0.0001)	0.44	(0.17)	0.00	(0.09) NS	0.00	(0.04) NS
ABSI	0.0023	(0.0007)	0.0042	(0.0001)	0.36	(0.17)	0.00	(0.08) NS	0.00	(0.04) NS
Urea	0.0421	(0.0100)	0.0565	(0.0008)	0.12	(0.17) NS	0.09	(0.09) NS	0.01	(0.05) NS
Creatinine	0.0023	(0.0043) NS	0.0242	(0.0004)	0.68	(0.17)	0.00	(0.09) NS	0.08	(0.04)
Glucose	0.0090	(0.0023)	0.0131	(0.0002)	0.17	(0.17) NS	0.00	(0.09) NS	0.09	(0.05) NS
TC	0.0204	(0.0069)	0.0381	(0.0006)	0.28	(0.18)	0.04	(0.09) NS	0.11	(0.04)
HDL	0.0095	(0.0106) NS	0.0634	(0.0010)	0.66	(0.16)	0.00	(0.08) NS	0.01	(0.04) NS
SBP	0.0067	(0.0026)	0.0142	(0.0002)	0.33	(0.18) NS	0.00	(0.09) NS	0.07	(0.05) NS
DBP	0.0112	(0.0026)	0.0145	(0.0002)	0.10	(0.17) NS	0.05	(0.09) NS	0.03	(0.05) NS
HR	-	-	-	-	-	-	-	-	-	-

[f] This column shows the residual variance

[g] This column shows the total phenotypic variance

[h] This column shows the proportion of total phenotypic variance captured by matrix **GRM_{kin}**

[i] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Family}**

[j] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Sib}**

[p] GCTA automatically stops the analysis because more than half the variance components are constrained to 0.

Trait	Model: KFSC						
	e^2_c (s.e.)[k]		h^2_{gkin} (s.e.)[l]		%V_C[m]	logL[n]	n[o]
Height	0.17	(0.09) [NS]	0.82	(0.17)	100.00%	-20956.27	9,150
Weight	0.08	(0.09) [NS]	0.44	(0.17)	62.49%	10911.82	9,118
Fat	0.17	(0.09)	0.48	(0.17)	68.91%	-21500.61	8,926
BMI	0.13	(0.09) [NS]	0.43	(0.17)	64.45%	11515.82	9,107
Hips	0.12	(0.09) [NS]	0.39	(0.18)	55.58%	16791.95	8,984
Waist	0.24	(0.09) [NS]	0.61	(0.17)	84.25%	13524.41	9,016
WHR	0.12	(0.09) [NS]	0.44	(0.17)	56.22%	18146.24	8,995
ABSI	0.08	(0.09) [NS]	0.36	(0.17)	43.64%	19937.55	8,962
Urea	0.04	(0.09) [NS]	0.12	(0.17) [NS]	25.31%	8520.81	9,148
Creatinine	0.15	(0.09) [NS]	0.68	(0.17)	90.72%	12532.22	9,146
Glucose	0.05	(0.09) [NS]	0.17	(0.17) [NS]	31.34%	14809.73	8,936
TC	0.03	(0.09) [NS]	0.28	(0.18)	46.64%	10324.79	9,136
HDL	0.19	(0.08) [NS]	0.66	(0.16)	85.01%	8040.84	9,125
SBP	0.13	(0.09) [NS]	0.33	(0.18) [NS]	52.24%	14807.32	9,144
DBP	0.04	(0.09) [NS]	0.10	(0.17) [NS]	22.27%	14699.52	9,141
HR	-	-	-	-	-	-	-

[k] This column shows the proportion of total phenotypic variance captured by matrix **ERM_{Couple}**

[l] This column shows the total heritability estimate, which is the sum of h^2_g and h^2_{kin}

[m] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[n] This column shows the log likelihood ratio for each analysis

[o] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

Table S2.4 (31) Results of variance component analysis using alternative model for 16 traits in GS10K: model 'GKFSC'

Trait	Model: GKFSC[a]							
	σ^2_g (s.e.)[b]		σ^2_{kin} (s.e.)[c]		σ^2_{ef} (s.e.)[d]		σ^2_{es} (s.e.)[e]	
Height	18.7138	(1.7485)	15.0836	(6.9909) NS	0.0001	(3.4552) NS	0.6869	(1.4458) NS
Weight	0.0095	(0.0014)	0.0056	(0.0059) NS	0.0033	(0.0029) NS	0.0002	(0.0012) NS
Fat	12.3709	(1.9587)	9.8925	(8.2477) NS	1.0001	(4.0946) NS	1.0929	(1.9107) NS
BMI	0.0075	(0.0012)	0.0058	(0.0052) NS	0.0021	(0.0026) NS	0.0000	(0.0011) NS
Hips	0.0018	(0.0004)	0.0015	(0.0016) NS	0.0004	(0.0008) NS	0.0000	(0.0003) NS
Waist	0.0027	(0.0007)	0.0080	(0.0032) NS	0.0000	(0.0016) NS	0.0000	(0.0007) NS
WHR	0.0008	(0.0003)	0.0018	(0.0011) NS	0.0000	(0.0006) NS	0.0000	(0.0003) NS
ABSI	0.0003	(0.0002)	0.0011	(0.0007) NS	0.0000	(0.0004) NS	0.0000	(0.0002) NS
Urea	0.0085	(0.0022)	0.0000	(0.0098) NS	0.0045	(0.0050) NS	0.0002	(0.0026) NS
Creatinine	0.0059	(0.0010)	0.0094	(0.0043)	0.0007	(0.0021) NS	0.0017	(0.0009)
Glucose	0.0025	(0.0005)	0.0000	(0.0022) NS	0.0000	(0.0011) NS	0.0011	(0.0007) NS
TC	0.0058	(0.0015)	0.0044	(0.0068) NS	0.0018	(0.0034) NS	0.0044	(0.0016)
HDL	0.0183	(0.0026)	0.0224	(0.0105) NS	0.0000	(0.0052) NS	0.0005	(0.0022) NS
SBP	0.0020	(0.0006)	0.0025	(0.0026) NS	0.0000	(0.0013) NS	0.0011	(0.0006) NS
DBP	0.0019	(0.0006)	0.0000	(0.0025) NS	0.0006	(0.0013) NS	0.0004	(0.0007) NS
HR	*0.0007*	*(0.0010)*[r]	0.0000	(0.0047) NS	0.0025	(0.0024) NS	0.0000	(0.0013) NS

[a] Model '**GKFSC**' = $\mathbf{GRM_g}$ +$\mathbf{GRM_{kin}}$ +$\mathbf{ERM_{Family}}$ + $\mathbf{ERM_{Sib}}$ + $\mathbf{ERM_{Couple}}$

[b] This column shows the variance captured by matrix $\mathbf{GRM_g}$

[c] This column shows the variance captured by matrix $\mathbf{GRM_{kin}}$

[d] This column shows the variance captured by matrix $\mathbf{ERM_{Family}}$

[e] This column shows the variance captured by matrix $\mathbf{ERM_{Sib}}$

Trait	Model: GKFSC									
	σ^2_{ec} (s.e.)[f]		σ^2_{ε} (s.e.)[g]		V (s.e.)[h]		h^2_g (s.e.)[i]		h^2_{kin} (s.e.)[j]	
Height	6.9224	(3.5493) [NS]	0.0001	(7.0273)	41.4069	(0.6673)	0.45	(0.04)	0.36	(0.17) [NS]
Weight	0.0029	(0.0030) [NS]	0.0125	(0.0059)	0.0341	(0.0005)	0.28	(0.04)	0.17	(0.17) [NS]
Fat	7.7178	(4.2284)	14.8906	(8.3295)	46.9647	(0.7269)	0.26	(0.04)	0.21	(0.18) [NS]
BMI	0.0041	(0.0026)	0.0101	(0.0052)	0.0296	(0.0005)	0.25	(0.04)	0.19	(0.17) [NS]
Hips	0.0010	(0.0008) [NS]	0.0039	(0.0016)	0.0087	(0.0001)	0.21	(0.04)	0.18	(0.18) [NS]
Waist	0.0044	(0.0016)	0.0033	(0.0032)	0.0183	(0.0003)	0.15	(0.04)	0.44	(0.17) [NS]
WHR	0.0008	(0.0006) [NS]	0.0029	(0.0011)	0.0064	(0.0001)	0.13	(0.04)	0.29	(0.17) [NS]
ABSI	0.0004	(0.0004) [NS]	0.0024	(0.0007)	0.0042	(0.0001)	0.08	(0.04)	0.27	(0.17) [NS]
Urea	0.0021	(0.0051)	0.0416	(0.0101)	0.0569	(0.0009)	0.15	(0.04)	0.00	(0.17) [NS]
Creatinine	0.0031	(0.0021) [NS]	0.0034	(0.0043)	0.0242	(0.0004)	0.24	(0.04)	0.39	(0.18)
Glucose	0.0007	(0.0012)	0.0088	(0.0023)	0.0131	(0.0002)	0.19	(0.04)	0.00	(0.17) [NS]
TC	0.0010	(0.0035) [NS]	0.0207	(0.0069)	0.0381	(0.0006)	0.15	(0.04)	0.12	(0.18) [NS]
HDL	0.0121	(0.0053) [NS]	0.0102	(0.0106)	0.0635	(0.0010)	0.29	(0.04)	0.35	(0.16) [NS]
SBP	0.0018	(0.0013) [NS]	0.0067	(0.0026)	0.0142	(0.0002)	0.14	(0.04)	0.18	(0.18) [NS]
DBP	0.0005	(0.0013) [NS]	0.0111	(0.0026)	0.0145	(0.0002)	0.13	(0.04)	0.00	(0.18) [NS]
HR	0.0000	(0.0025) [NS]	0.0226	(0.0049)	0.0258	(0.0004)	*0.03*	*(0.04)*	0.00	(0.18) [NS]

[f] This column shows the variance captured by matrix **ERM_{Couple}**

[g] This column shows the residual variance

[h] This column shows the total phenotypic variance

[i] This column shows the proportion of total phenotypic variance captured by matrix **GRM_g**

[j] This column shows the proportion of total phenotypic variance captured by matrix **GRM_{kin}**

Trait	Model: GKFSC							
	e^2_f (s.e.)[k]		e^2_s (s.e.)[l]		e^2_c (s.e.)[m]		h^2_{gkin} (s.e.)[n]	
Height	0.00	(0.08) NS	0.02	(0.03) NS	0.17	(0.09) NS	0.81	(0.12)
Weight	0.10	(0.09) NS	0.01	(0.04) NS	0.09	(0.09) NS	0.45	(0.12)
Fat	0.02	(0.09) NS	0.02	(0.04) NS	0.16	(0.09)	0.47	(0.13)
BMI	0.07	(0.09) NS	0.00	(0.04) NS	0.14	(0.09)	0.44	(0.12)
Hips	0.05	(0.09) NS	0.00	(0.04) NS	0.12	(0.09) NS	0.39	(0.13)
Waist	0.00	(0.09) NS	0.00	(0.04) NS	0.24	(0.09)	0.59	(0.12)
WHR	0.00	(0.09) NS	0.00	(0.04) NS	0.13	(0.09) NS	0.42	(0.12)
ABSI	0.00	(0.08) NS	0.00	(0.04) NS	0.08	(0.09) NS	0.35	(0.12)
Urea	0.08	(0.09) NS	0.00	(0.05) NS	0.04	(0.09)	0.15	(0.12)
Creatinine	0.03	(0.09) NS	0.07	(0.04)	0.13	(0.09) NS	0.63	(0.13)
Glucose	0.00	(0.09) NS	0.09	(0.05) NS	0.05	(0.09)	0.19	(0.12)
TC	0.05	(0.09) NS	0.12	(0.04)	0.02	(0.09) NS	0.27	(0.13)
HDL	0.00	(0.08) NS	0.01	(0.04) NS	0.19	(0.08) NS	0.64	(0.12)
SBP	0.00	(0.09) NS	0.08	(0.05) NS	0.13	(0.09) NS	0.32	(0.13)
DBP	0.04	(0.09) NS	0.03	(0.05) NS	0.03	(0.09) NS	0.13	(0.13)
HR	0.10	(0.09) NS	0.00	(0.05) NS	0.00	(0.10) NS	0.03	(0.13)

[k] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Family}}$

[l] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Sib}}$

[m] This column shows the proportion of total phenotypic variance captured by matrix $\mathbf{ERM_{Couple}}$

[n] This column shows the total heritability estimate, which is the sum of h^2_g and h^2_{kin}

Trait	Model: GKFSC		
	%V_C[o]	logL[p]	n[q]
Height	100.00%	-20836.73	9,150
Weight	63.02%	10937.32	9,118
Fat	68.29%	-21479.20	8,926
BMI	65.90%	11537.52	9,107
Hips	54.29%	16806.26	8,984
Waist	82.31%	13533.89	9,016
WHR	53.67%	18154.22	8,995
ABSI	43.41%	19941.09	8,962
Urea	26.86%	8529.10	9,148
Creatinine	85.90%	12551.94	9,146
Glucose	32.72%	14822.23	8,936
TC	45.87%	10332.99	9,136
HDL	83.97%	8071.55	9,125
SBP	52.18%	14814.96	9,144
DBP	23.46%	14705.67	9,141
HR	12.45%	12056.49	9,126

[o] This column shows the proportion of total phenotypic variance captured by this model for each analysis

[p] This column shows the log likelihood ratio for each analysis

[q] This column shows the number of records used for each analysis

[NS] Not significant. The variance component is not significantly different from 0 (log likelihood test, P-value > 0.05)

[r] Italic means we are not able to get the P-value for the log likelihood ratio test for this variance component since GCTA abrogates the analysis once we drop this variance component out.

Table S2.5

Table S2.5 Parameter settings and main results (mean and 95% CI) for all scenarios in simulation study using GS10K data.

Scenario	Phenotype contributors					Model	Samples	No. Replicates
	h^2_g	h^2_{kin}	e^2_f	e^2_s	e^2_c			
i	h^2_g= 0.2 / 0.3 / 0.5	h^2_{kin}= 0	e^2_f= 0	e^2_s= 0	e^2_c= 0	G	ALL	50
	h^2_g= 0	h^2_{kin}= 0.2 / 0.3 / 0.5	e^2_f= 0	e^2_s= 0	e^2_c= 0	K		
	h^2_g= 0	h^2_{kin}= 0	e^2_f= 0.015 / 0.025 / 0.05 / 0.1	e^2_s= 0	e^2_c= 0	F		
	h^2_g= 0	h^2_{kin}= 0	e^2_f= 0	e^2_s= 0.03 / 0.05 / 0.1 / 0.2	e^2_c= 0	S		
	h^2_g= 0	h^2_{kin}= 0	e^2_f= 0	e^2_s= 0	e^2_c= 0.03 / 0.05 / 0.1 / 0.2	C		

Scenario	Mean of overall estimates (95% CI)		
	h^2_g	h^2_{kin}	e^2_f
i	h^2_g= 0.203 (0.196-0.209) h^2_g= **0.311 (0.305-0.317)**[a] h^2_g= **0.508 (0.501-0.515)**	h^2_{kin}= -	e^2_f= -
	h^2_g= -	h^2_{kin}= 0.203 (0.194-0.213) h^2_{kin}= 0.296 (0.287-0.305) h^2_{kin}= 0.499 (0.489-0.508)	e^2_f= -
	h^2_g= -	h^2_{kin}= -	e^2_f= 0.016 (0.012-0.020) e^2_f= 0.027 (0.024-0.030) e^2_f= 0.050 (0.046-0.053) e^2_f= 0.099 (0.094-0.104)
	h^2_g= -	h^2_{kin}= -	e^2_f= -
	h^2_g= -	h^2_{kin}= -	e^2_f= -

[a] Values in bold means that they are significantly different from parameter settings according to Z-test at $\alpha = 0.05$ level.

Scenario	Mean of overall estimates (95% CI)	
	e^2_s	e^2_c
i	e^2_s= -	e^2_c= -
	e^2_s= -	e^2_c= -
	e^2_s= -	e^2_c= -
	e^2_s= 0.031 (0.023-0.038)	e^2_c= -
	e^2_s= 0.057 (0.046-0.067)	
	e^2_s= 0.092 (0.081-0.103)	
	e^2_s= 0.200 (0.188-0.212)	
	e^2_s= -	e^2_c= 0.035 (0.028-0.043)
		e^2_c= 0.053 (0.045-0.060)
		e^2_c= 0.097 (0.090-0.104)
		e^2_c= 0.199 (0.191-0.206)

Scenario	Phenotype contributors										Model	Samples	No. Replicates
	h^2_g		h^2_{kin}		e^2_f		e^2_s		e^2_c				
ii	h^2_g=	0.5	h^2_{kin}=	0	e^2_f=	0	e^2_s=	0	e^2_c=	0	GK	All	50
	h^2_g=	0.5	h^2_{kin}=	0.2									
	h^2_g=	0.3	h^2_{kin}=	0.3									
	h^2_g=	0.2	h^2_{kin}=	0.5									
	h^2_g=	0.3	h^2_{kin}=	0	e^2_f=	0.05	e^2_s=	0	e^2_c=	0	GF		
	h^2_g=	0.5	h^2_{kin}=	0.2	e^2_f=	0	e^2_s=	0	e^2_c=	0	G		
	h^2_g=	0.3	h^2_{kin}=	0	e^2_f=	0.05	e^2_s=	0	e^2_c=	0	G		
	h^2_g=	0.3	h^2_{kin}=	0.2	e^2_f=	0	e^2_s=	0	e^2_c=	0	GF		
	h^2_g=	0.3	h^2_{kin}=	0	e^2_f=	0.05	e^2_s=	0	e^2_c=	0	GK		
	h^2_g=	0.5	h^2_{kin}=	0.2	e^2_f=	0	e^2_s=	0	e^2_c=	0	G	Unrelated Individuals	
	h^2_g=	0.3	h^2_{kin}=	0	e^2_f=	0.05	e^2_s=	0	e^2_c=	0	G		
iii	h^2_g=	0.3	h^2_{kin}=	0	e^2_f=	0.05	e^2_s=	0.1	e^2_c=	0	GFS	ALL	100
	h^2_g=	0.3	h^2_{kin}=	0.2	e^2_f=	0	e^2_s=	0	e^2_c=	0.1	GKC		
	h^2_g=	0.3	h^2_{kin}=	0.2	e^2_f=	0	e^2_s=	0.1	e^2_c=	0.1	GKSC		
	h^2_g=	0.3	h^2_{kin}=	0.2	e^2_f=	0.05	e^2_s=	0.1	e^2_c=	0.1	GKFSC		

Scenario	Mean of overall estimates (95% CI)					
	h^2_g		h^2_{kin}		e^2_f	
ii	h^2_g=	0.502 (0.495-0.509)	h^2_{kin}=	**0.008 (0.003-0.013)**	e^2_f=	-
		0.525 (0.516-0.534)		**0.169 (0.160-0.181)**		
		0.318 (0.309-0.326)		**0.277 (0.267-0.287)**		
		0.214 (0.205-0.223)		**0.480 (0.468-0.491)**		
	h^2_g=	**0.309 (0.303-0.315)**	h^2_{kin}=	-	e^2_f=	**0.044 (0.040-0.048)**
	h^2_g=	**0.608 (0.602-0.615)**	h^2_{kin}=	-	e^2_f=	-
	h^2_g=	**0.341 (0.334-0.348)**	h^2_{kin}=	-	e^2_f=	-
	h^2_g=	**0.350 (0.342-0.359)**	h^2_{kin}=	-	e^2_f=	**0.060 (0.055-0.066)**
	h^2_g=	**0.311 (0.302-0.320)**	h^2_{kin}=	**0.072 (0.061-0.082)**	e^2_f=	-
	h^2_g=	**0.526 (0.515-0.537)**	h^2_{kin}=	-	e^2_f=	-
	h^2_g=	0.310 (0.297-0.323)	h^2_{kin}=	-	e^2_f=	-
iii	h^2_g=	**0.308 (0.302-0.314)**	h^2_{kin}=	-	e^2_f=	**0.044 (0.040-0.048)**
	h^2_g=	**0.311 (0.305-0.317)**	h^2_{kin}=	**0.188 (0.179-0.197)**	e^2_f=	-
	h^2_g=	**0.310 (0.304-0.316)**	h^2_{kin}=	**0.190 (0.182-0.198)**	e^2_f=	-
	h^2_g=	0.304 (0.298-0.310)	h^2_{kin}=	0.180 (0.156-0.204)	e^2_f=	**0.063 (0.052-0.073)**

Scenario	Mean of overall estimates (95% CI)	
	e^2_s	e^2_c
ii	e^2_s= -	e^2_c= -
	e^2_s= -	e^2_c= -
	e^2_s= -	e^2_c= -
	e^2_s= -	e^2_c= -
	e^2_s= -	e^2_c= -
	e^2_s= -	e^2_c= -
	e^2_s= -	e^2_c= -
	e^2_s= -	e^2_c= -
iii	e^2_s= 0.105 (0.097-0.112)	e^2_c= -
	e^2_s= -	e^2_c= 0.099 (0.094-0.105)
	e^2_s= 0.101 (0.095-0.107)	e^2_c= 0.099 (0.094-0.104)
	e^2_s= 0.097 (0.091-0.103)	e^2_c= 0.093 (0.081-0.105)

Table S2.6

Table S2.6 Parameter settings for phenotypes simulated using GS10K data, results of model selection using simulated phenotypes and results of variance component analyses using selected models.

Scenario	Phenotype Contributors					No.	Model	Estimates of effects (s.e.)									
	h^2_g	h^2_g	e^2_f	e^2_s	e^2_c			h^2_g		h^2_g		e^2_f		e^2_s		e^2_c	
a	h^2_g=0.2	h^2_{kin}=0.05	e^2_f=0.15	e^2_s=0.03	e^2_c=0.05	1	GF	h^2_g =	0.21 (0.02)	h^2_{kin} =	-	e^2_f =	0.15 (0.02)	e^2_s =	-	e^2_c =	-
						2	GF	h^2_g =	0.25 (0.03)	h^2_{kin} =	-	e^2_f =	0.13 (0.02)	e^2_s =	-	e^2_c =	-
						3	GF	h^2_g =	0.16 (0.02)	h^2_{kin} =	-	e^2_f =	0.18 (0.02)	e^2_s =	-	e^2_c =	-
						4	GFC	h^2_g =	0.24 (0.03)	h^2_{kin} =	-	e^2_f =	0.16 (0.02)	e^2_s =	-	e^2_c =	0.07 (0.03)
						5	GFS	h^2_g =	0.20 (0.02)	h^2_{kin} =	-	e^2_f =	0.15 (0.02)	e^2_s =	0.10 (0.03)	e^2_c =	-
						6	GF	h^2_g =	0.23 (0.02)	h^2_{kin} =	-	e^2_f =	0.17 (0.02)	e^2_s =	-	e^2_c =	-
						7	GKC	h^2_g =	0.23 (0.03)	h^2_{kin} =	0.38 (0.04)	e^2_f =	-	e^2_s =	-	e^2_c =	0.23 (0.02)
						8	GF	h^2_g =	0.21 (0.02)	h^2_{kin} =	-	e^2_f =	0.18 (0.02)	e^2_s =	-	e^2_c =	-
						9	GFC	h^2_g =	0.24 (0.03)	h^2_{kin} =	-	e^2_f =	0.13 (0.02)	e^2_s =	-	e^2_c =	0.09 (0.03)
						10	GF	h^2_g =	0.19 (0.02)	h^2_{kin} =	-	e^2_f =	0.20 (0.02)	e^2_s =	-	e^2_c =	-

Scenario	Phenotype Contributors					No.	Model	Estimates of effects (s.e.)									
	h^2_g	h^2_g	e^2_f	e^2_s	e^2_c			h^2_g		h^2_g		e^2_f		e^2_s		e^2_c	
b	h^2_g=0.2	h^2_{kin}=0.3	e^2_f=0.025	e^2_s=0.03	e^2_c=0.2	1	GKC	h^2_g =	0.21 (0.03)	h^2_{kin} =	0.36(0.04)	e^2_f =	-	e^2_s =	-	e^2_c =	0.26(0.02)
						2	GFSC	h^2_g =	0.20 (0.03)	h^2_{kin} =	-	e^2_f =	0.18 (0.02)	e^2_s =	0.07 (0.03)	e^2_c =	0.09 (0.03)
						3	GKC	h^2_g =	0.21 (0.03)	h^2_{kin} =	0.40 (0.04)	e^2_f =	-	e^2_s =	-	e^2_c =	0.21 (0.03)
						4	GKSC	h^2_g =	0.19 (0.03)	h^2_{kin} =	0.40 (0.05)	e^2_f =	-	e^2_s =	0.06 (0.03)	e^2_c =	0.24 (0.03)
						5	GKC	h^2_g =	0.19 (0.03)	h^2_{kin} =	0.33 (0.04)	e^2_f =	-	e^2_s =	-	e^2_c =	0.23 (0.03)
						6	GKC	h^2_g =	0.18 (0.03)	h^2_{kin} =	0.38 (0.04)	e^2_f =	-	e^2_s =	-	e^2_c =	0.23 (0.03)
						7	GKC	h^2_g =	0.20 (0.03)	h^2_{kin} =	0.30 (0.04)	e^2_f =	-	e^2_s =	-	e^2_c =	0.18 (0.03)
						8	GKC	h^2_g =	0.22 (0.03)	h^2_{kin} =	0.33 (0.04)	e^2_f =	-	e^2_s =	-	e^2_c =	0.20 (0.02)
						9	GKSC	h^2_g =	0.18 (0.03)	h^2_{kin} =	0.35 (0.05)	e^2_f =	-	e^2_s =	0.06 (0.03)	e^2_c =	0.20 (0.03)
						10	GF	h^2_g =	0.17 (0.03)	h^2_{kin} =	-	e^2_f =	0.23 (0.02)	e^2_s =	-	e^2_c =	-

Scenario	Phenotype Contributors					No.	Model	Estimates of effects (s.e.)									
	h^2_g	h^2_g	e^2_f	e^2_s	e^2_c			h^2_g		h^2_g		e^2_f		e^2_s		e^2_c	
c	h^2_g=0.25	h^2_{kin}=0.15	e^2_f=0.15	e^2_s=0.05	e^2_c=0.15	1	GFSC	h^2_g =	0.27 (0.03)	h^2_{kin} =	-	e^2_f =	0.14 (0.02)	e^2_s =	0.05 (0.03)	e^2_c =	0.16 (0.03)
						2	GKFSC	h^2_g =	0.28 (0.03)	h^2_{kin} =	0.22 (0.14)	e^2_f =	0.11 (0.07)	e^2_s =	0.08 (0.03)	e^2_c =	0.20 (0.07)
						3	GKFSC	h^2_g =	0.28 (0.03)	h^2_{kin} =	0.13 (0.14)	e^2_f =	0.14 (0.07)	e^2_s =	0.03 (0.03)	e^2_c =	0.14 (0.07)
						4	GFSC	h^2_g =	0.22 (0.03)	h^2_{kin} =	-	e^2_f =	0.19 (0.02)	e^2_s =	0.04 (0.03)	e^2_c =	0.14 (0.03)
						5	GKFC	h^2_g =	0.24 (0.03)	h^2_{kin} =	0.17 (0.14)	e^2_f =	0.13 (0.07)	e^2_s =	-	e^2_c =	0.14 (0.07)
						6	GKFSC	h^2_g =	0.27 (0.03)	h^2_{kin} =	0.03 (0.14)	e^2_f =	0.18 (0.07)	e^2_s =	0.08 (0.03)	e^2_c =	0.12 (0.07)
						7	GKFSC	h^2_g =	0.31 (0.03)	h^2_{kin} =	0.12 (0.14)	e^2_f =	0.10 (0.07)	e^2_s =	0.09 (0.03)	e^2_c =	0.20 (0.07)
						8	GKFC	h^2_g =	0.29 (0.03)	h^2_{kin} =	0.05 (0.14)	e^2_f =	0.16 (0.07)	e^2_s =	-	e^2_c =	0.15 (0.07)
						9	GKFSC	h^2_g =	0.24 (0.03)	h^2_{kin} =	0.32 (0.14)	e^2_f =	0.11 (0.07)	e^2_s =	0.03 (0.03)	e^2_c =	0.18 (0.07)
						10	GKFC	h^2_g =	0.20 (0.03)	h^2_{kin} =	0.42 (0.14)	e^2_f =	0.06 (0.07)	e^2_s =	-	e^2_c =	0.20 (0.07)

Table S2.7

Table S2.7 Phenotypic correlation for spousal pairs (covariates adjusted) in GS10K

Trait	Correlation	95% Confidence Interval
Height	0.257	0.203 - 0.310
Weight	0.178	0.122 - 0.234
Fat	0.170	0.112 - 0.227
BMI	0.206	0.150 - 0.261
Hips	0.159	0.102 - 0.216
Waist	0.194	0.138 - 0.250
WHR	0.081	0.022 - 0.138
ABSI	0.048	-0.011 - 0.106[NS]
Urea	0.121	0.063 - 0.177
Creatinine	0.138	0.081 - 0.194
Glucose	0.055	-0.004 - 0.114[NS]
TC	0.063	0.006 - 0.120
HDL	0.147	0.090 - 0.203
SBP	0.107	0.050 - 0.164
DBP	0.091	0.033 - 0.148
HR	0.091	0.033 - 0.148

[NS] Not significant. The correlation is not significantly different from 0

Table S2.8

Table S2.8 Variance component analysis results of the full model and the selected models for depression, cognition and personality traits in GS:SFHS.

Phenotype	*N*	*Model*		GRM_g h^2_g (S.E)	GRM_{kin} h^2_{kin} (S.E)	ERM_{Family} e_f^2 (S.E)	ERM_{Sib} e_s^2 (S.E)	ERM_{Couple} e_c^2 (S.E)
Cognitive								
g	19 036	Full	GKFSC	0.21 (0.02)	0.42 (0.05)	0.00 (0.02)	0.09 (0.01)	0.26 (0.03)
		Selected	GKSC	0.23 (0.02)	0.31 (0.03)		0.09 (0.01)	0.22 (0.02)
Education	18 528	Full	GKFSC	0.13 (0.02)	0.39 (0.05)	0.00 (0.02)	0.11 (0.01)	0.36 (0.03)
		Selected	GKSC	0.16 (0.02)	0.28 (0.03)		0.11 (0.01)	0.31 (0.03)
Vocabulary	19 269	Full	GKFSC	0.23 (0.02)	0.39 (0.05)	0.00 (0.02)	0.07 (0.01)	0.31 (0.03)
		Selected	GKSC	0.26 (0.02)	0.30 (0.03)		0.07 (0.01)	0.27 (0.02)
Verbal Fluency	19 380	Full	GKFSC	0.18 (0.02)	0.31 (0.05)	0.00 (0.02)	0.05 (0.01)	0.16 (0.03)
		Selected	GKSC	0.19 (0.02)	0.27 (0.03)		0.05 (0.01)	0.15 (0.02)
Digit Symbol Test	19 385	Full	GKFSC	0.20 (0.02)	0.23 (0.05)	0.00 (0.02)	0.08 (0.01)	0.17 (0.03)
		Selected	GKSC	0.21 (0.02)	0.15 (0.03)		0.08 (0.01)	0.13 (0.02)
Logical Memory	19 365	Full	GKFSC	0.11 (0.02)	0.24 (0.05)	0.00 (0.02)	0.05 (0.01)	0.05 (0.03)
		Selected	GKS	0.12 (0.02)	0.20 (0.03)		0.05 (0.01)	
Personality								
Neuroticism	19 494	Full	GKFSC	0.11 (0.02)	0.15 (0.05)	0.02 (0.03)	0.00 (0.01)	0.01 (0.03)
		Selected	GK	0.11 (0.02)	0.19 (0.03)			
Extraversion	19 487	Full	GKFSC	0.11 (0.02)	0.05 (0.05)	0.07 (0.03)	0.00 (0.01)	0.01 (0.03)
		Selected	GF	0.13 (0.02)		0.09 (0.01)		
Depression								
Major Depressive Disorder	19 896	Full	GKFSC	0.10 (0.05)	0.20 (0.12)	0.09 (0.06)	0.00 (0.04)	0.03 (0.09)
		Selected	GKC	0.12 (0.05)	0.35 (0.06)			0.14 (0.07)

Table S3.1

Table S3.1 Number of unique and common signals detected by each method.

Trait	Common			Unique		
	ALL	TU & TR/SR	TR & SR	TU	TR	SR
Glucose	10	0	30	3	0	1
Waist	2	0	5	2	2	3
Fat	3	0	7	2	2	4
HDL_C	29	0	46	5	4	8
Weight	3	0	8	5	2	4
Height	5	0	40	9	21	6
Hip	2	0	7	5	3	3
DBP	0	0	2	4	0	0
WHR	0	0	4	4	1	1
ABSI	0	0	4	4	0	3
HR	1	0	8	2	5	2
BMI	5	0	10	0	2	2
SBP	0	0	11	9	0	2
TC	18	0	50	2	0	5
Creatinine	0	0	19	6	5	6
Urea	0	0	16	2	1	1

Definition of unique and common signals see Chapter 3: 3.3.1. ALL means a signal can be detected by all three methods. TU & TR/SR means a signal can be detected by both TU and TR methods or by both TU and SR methods. TR & SR means a signal can be detected by both TR and SR methods

Table S3.2

Table S3.2 Number of suggestive trait-associated SNPs detected by each method having strong, some and no supporting evidence.

Trait	Common TU TR SR			Common TR SR			Unique TU		
	No Evidence	Some Evidence	Strong Evidence	No Evidence	Some Evidence	Strong Evidence	No Evidence	Some Evidence	Strong Evidence
Glucose	1	4	5	9	5	16	3	0	0
Waist	0	0	2	3	0	2	2	0	0
Fat	0	0	3	6	0	1	2	0	0
HDL_C	0	1	28	2	11	33	4	0	1
Weight	0	3	0	4	4	0	5	0	0
Height	0	0	5	1	2	37	3	0	6
Hip	0	1	1	5	0	2	5	0	0
DBP	0	0	0	2	0	0	4	0	0
WHR	0	0	0	4	0	0	3	1	0
HR	1	0	0	2	0	6	1	0	1
BMI	0	0	5	7	0	3	0	0	0
SBP	0	0	0	9	2	0	9	0	0
TC	0	1	17	6	9	35	2	0	0
Creatinine	0	0	0	9	1	9	6	0	0
Urea	0	0	0	10	2	4	2	0	0
Sum	2	10	66 (61)	79 (78)	36 (34)	148 (111)	51 (48)	1	8 (2)
Pw/oE	2.56% (2.74%)			30.04% (34.98%)			85.00% (94.12%)		

Trait	Unique TR			Unique SR			TU		
	No Evidence	Some Evidence	Strong Evidence	No Evidence	Some Evidence	Strong Evidence	No Evidence	Some Evidence	Strong Evidence
Glucose	0	0	0	0	0	1	4	4	5
Waist	2	0	0	3	0	0	2	0	2
Fat	2	0	0	4	0	0	2	0	3
HDL_C	2	1	1	0	4	4	4	1	29
Weight	2	0	0	4	0	0	5	3	0
Height	2	0	19	0	0	6	3	0	11
Hip	3	0	0	3	0	0	5	1	1
DBP	0	0	0	0	0	0	4	0	0
WHR	1	0	0	1	0	0	3	1	0
HR	4	0	1	0	0	2	2	0	1
BMI	2	0	0	0	0	2	0	0	5
SBP	0	0	0	1	0	1	9	0	0
TC	0	0	0	0	2	3	2	1	17
Creatinine	3	1	1	3	2	1	6	0	0
Urea	0	0	1	1	0	0	2	0	0
Sum	23 (21)	2	23 (4)	20	8	20 (14)	53 (50)	11	74 (63)
Pw/oE	47.92% (77.78%)			41.67% (47.62%)			38.41% (40.32%)		

Trait	TR			SR		
	No Evidence	Some Evidence	Strong Evidence	No Evidence	Some Evidence	Strong Evidence
Glucose	10	9	21	10	9	22
Waist	5	0	4	6	0	4
Fat	8	0	4	10	0	4
HDL_C	4	13	62	2	16	65
Weight	6	7	0	8	7	0
Height	3	2	61	1	2	48
Hip	8	1	3	8	1	3
DBP	2	0	0	2	0	0
WHR	5	0	0	5	0	0
HR	7	0	7	3	0	8
BMI	9	0	8	7	0	10
SBP	9	2	0	10	2	1
TC	6	10	52	6	12	55
Creatinine	12	2	10	12	3	10
Urea	10	2	5	11	2	4
Sum	104 (101)	48 (46)	237 (176)	101 (100)	54 (52)	234 (186)
Pw/oE	26.74% (31.27%)			25.96% (29.59%)		

Pw/oE: the proportion of SNPs without evidence, i.e. No evidence / (No evidence + some evidence + strong evidence)

The values in bracket () are the results excluding height.

Table S4.1

Table S4.1. Results of 5-fold KRR prediction analysis using alternative models for height in GS10K.

Lambda	0.0001		0.001		0.01		1		10		100	
Model	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE
G	35.75%	0.8871	35.84%	0.8847	36.47%	0.8696	37.50%	0.8705	36.52%	0.9576	35.82%	0.9917
K	30.52%	0.9065	30.58%	0.9057	31.00%	0.9009	31.95%	0.9133	31.67%	0.9739	31.44%	0.9938
F	2.04%	3951.5946	0.29%	165.1776	20.29%	1.3591	30.02%	0.9087	30.42%	0.9531	30.14%	0.9906
S	15.63%	0.9882	15.63%	0.9876	15.64%	0.9826	15.64%	0.9722	15.38%	0.9884	15.23%	0.9955
C	0.59%	16.0674	8.89%	1.1807	11.81%	1.0511	11.89%	0.9904	11.90%	0.9889	11.90%	0.9955
GK	36.80%	0.8635	36.87%	0.8626	37.35%	0.8575	38.15%	0.8786	37.34%	0.9637	36.89%	0.9925
GF	35.81%	0.8906	36.24%	0.8837	37.49%	0.8612	38.58%	0.8634	37.56%	0.9550	37.02%	0.9913
GS	36.87%	0.8639	36.90%	0.8632	37.16%	0.8591	37.54%	0.8770	36.45%	0.9615	35.82%	0.9922
GC	37.25%	0.8583	37.27%	0.8581	37.46%	0.8576	37.78%	0.8819	36.86%	0.9639	36.32%	0.9925
KF	31.41%	0.9092	31.50%	0.9074	32.08%	0.8970	32.75%	0.9004	32.04%	0.9672	31.69%	0.9929
KS	31.05%	0.9007	31.08%	0.9004	31.32%	0.8990	31.85%	0.9165	31.44%	0.9751	31.20%	0.9940
KC	32.95%	0.8917	32.95%	0.8919	32.95%	0.8944	32.75%	0.9198	32.07%	0.9767	31.85%	0.9942
FS	0.46%	7997.7906	5.78%	8.0589	24.43%	1.0620	30.19%	0.9058	30.42%	0.9574	30.14%	0.9913
FC	2.04%	3951.5946	0.29%	165.1776	20.29%	1.3591	30.02%	0.9087	30.42%	0.9531	30.14%	0.9906
SC	19.50%	0.9608	19.51%	0.9605	19.54%	0.9590	19.55%	0.9637	19.30%	0.9881	19.19%	0.9955
GKF	17.73%	1.6237	36.35%	0.8816	37.51%	0.8607	38.58%	0.8636	37.56%	0.9552	37.02%	0.9913
GKS	37.45%	0.8568	37.48%	0.8566	37.73%	0.8556	38.09%	0.8827	37.18%	0.9654	36.74%	0.9928
GKC	38.67%	0.8530	38.67%	0.8534	38.69%	0.8569	38.45%	0.8918	37.48%	0.9691	37.12%	0.9932
GFS	36.75%	0.8749	36.91%	0.8721	37.74%	0.8574	38.57%	0.8648	37.53%	0.9559	37.00%	0.9914
GFC	37.49%	0.8620	37.56%	0.8608	38.01%	0.8534	38.55%	0.8673	37.54%	0.9573	37.03%	0.9916
GSC	37.52%	0.8568	37.54%	0.8569	37.67%	0.8579	37.84%	0.8861	36.87%	0.9659	36.36%	0.9928
KFS	31.81%	0.9019	31.87%	0.9009	32.26%	0.8943	32.71%	0.9020	31.98%	0.9681	31.64%	0.9930

Lambda	500		1000	
Model	Cor	MSE	Cor	MSE
G	35.73%	0.9955	35.71%	0.9960
K	31.42%	0.9959	31.41%	0.9962
F	30.10%	0.9953	30.09%	0.9959
S	15.21%	0.9963	15.21%	0.9964
C	11.90%	0.9963	11.90%	0.9964
GK	36.84%	0.9957	36.83%	0.9961
GF	36.96%	0.9954	36.95%	0.9960
GS	35.74%	0.9956	35.73%	0.9961
GC	36.25%	0.9957	36.24%	0.9961
KF	31.65%	0.9958	31.65%	0.9961
KS	31.17%	0.9960	31.17%	0.9962
KC	31.82%	0.9960	31.82%	0.9963
FS	30.10%	0.9954	30.10%	0.9960
FC	30.10%	0.9953	30.09%	0.9959
SC	19.18%	0.9963	19.18%	0.9964
GKF	36.96%	0.9954	36.95%	0.9960
GKS	36.69%	0.9957	36.68%	0.9961
GKC	37.08%	0.9958	37.07%	0.9962
GFS	36.93%	0.9955	36.92%	0.9960
GFC	36.97%	0.9955	36.96%	0.9960
GSC	36.30%	0.9957	36.29%	0.9961
KFS	31.60%	0.9958	31.60%	0.9961

Lambda	0.0001		0.001		0.01		1		10		100	
Model	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE
KFC	32.92%	0.8921	32.92%	0.8923	32.92%	0.8949	32.70%	0.9203	32.04%	0.9768	31.82%	0.9942
KSC	32.92%	0.8923	32.92%	0.8926	32.92%	0.8952	32.70%	0.9207	32.02%	0.9770	31.79%	0.9942
FSC	0.11%	1043.4989	5.91%	11.4261	22.50%	1.1885	30.09%	0.9075	30.43%	0.9543	30.15%	0.9908
GKFS	36.82%	0.8735	36.97%	0.8710	37.76%	0.8570	38.58%	0.8651	37.53%	0.9561	37.00%	0.9915
GKFC	38.41%	0.8569	38.41%	0.8572	38.40%	0.8611	38.11%	0.8962	37.18%	0.9706	36.85%	0.9934
GKSC	38.68%	0.8536	38.68%	0.8539	38.69%	0.8576	38.43%	0.8928	37.45%	0.9694	37.09%	0.9933
GFSC	37.72%	0.8582	37.77%	0.8573	38.13%	0.8519	38.54%	0.8688	37.52%	0.9582	37.01%	0.9918
KFSC	32.90%	0.8925	32.90%	0.8927	32.90%	0.8953	32.68%	0.9208	32.02%	0.9770	31.79%	0.9942
GKFSC	38.38%	0.8574	38.38%	0.8578	38.37%	0.8617	38.07%	0.8968	37.14%	0.9708	36.82%	0.9935

Lambda	500		1000	
Model	Cor	MSE	Cor	MSE
KFC	31.79%	0.9960	31.79%	0.9963
KSC	31.77%	0.9960	31.76%	0.9963
FSC	30.11%	0.9953	30.10%	0.9959
GKFS	36.93%	0.9955	36.92%	0.9960
GKFC	36.82%	0.9959	36.81%	0.9962
GKSC	37.05%	0.9958	37.04%	0.9962
GFSC	36.95%	0.9955	36.94%	0.9960
KFSC	31.77%	0.9960	31.76%	0.9963
GKFSC	36.78%	0.9959	36.78%	0.9962

Cor: correlation; MSE: mean squared errors

Table S4.2

Table S4.2. Results of 5-fold KRR prediction analysis using alternative models for BMI in GS10K.

Lambda	0.0001		0.001		0.01		1		10		100	
Model	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE
G	16.90%	1.0739	16.98%	1.0688	17.54%	1.0329	18.81%	0.9641	18.87%	0.9812	18.66%	0.9951
K	16.38%	1.0083	16.44%	1.0062	16.90%	0.9909	18.19%	0.9643	18.55%	0.9855	18.52%	0.9958
F	0.98%	5070.1706	1.21%	392.2583	11.50%	1.5601	18.83%	0.9808	19.25%	0.9739	19.14%	0.9939
S	10.58%	1.0198	10.58%	1.0189	10.56%	1.0112	10.46%	0.9882	10.19%	0.9922	10.07%	0.9966
C	1.39%	35.9085	5.98%	1.3350	8.57%	1.0721	8.49%	1.0040	8.46%	0.9923	8.46%	0.9965
GK	17.97%	1.0173	18.04%	1.0145	18.55%	0.9941	19.82%	0.9580	20.02%	0.9827	19.90%	0.9954
GF	18.80%	1.0622	18.91%	1.0549	19.63%	1.0151	21.03%	0.9540	21.01%	0.9782	20.82%	0.9947
GS	18.21%	1.0171	18.24%	1.0149	18.50%	0.9977	19.20%	0.9607	19.09%	0.9825	18.87%	0.9953
GC	19.56%	0.9838	19.57%	0.9827	19.72%	0.9740	20.13%	0.9570	19.90%	0.9831	19.69%	0.9955
KF	17.85%	1.0213	17.92%	1.0181	18.48%	0.9961	19.57%	0.9592	19.55%	0.9816	19.43%	0.9952
KS	17.45%	0.9880	17.48%	0.9869	17.72%	0.9784	18.46%	0.9637	18.60%	0.9861	18.54%	0.9958
KC	19.74%	0.9617	19.74%	0.9615	19.79%	0.9596	19.85%	0.9613	19.62%	0.9866	19.53%	0.9959
FS	1.94%	3405.6768	3.41%	13.2289	13.99%	1.2891	19.10%	0.9711	19.33%	0.9760	19.21%	0.9943
FC	0.98%	5070.1706	1.21%	392.2583	11.50%	1.5601	18.83%	0.9808	19.25%	0.9739	19.14%	0.9939
SC	13.07%	0.9929	13.08%	0.9923	13.12%	0.9880	13.17%	0.9800	13.02%	0.9918	12.95%	0.9965
GKF	4.93%	3.8247	18.92%	1.0527	19.64%	1.0140	21.03%	0.9540	21.01%	0.9782	20.82%	0.9948
GKS	18.90%	0.9940	18.94%	0.9924	19.23%	0.9800	20.04%	0.9573	20.07%	0.9833	19.94%	0.9955
GKC	21.12%	0.9589	21.13%	0.9585	21.20%	0.9554	21.36%	0.9548	21.08%	0.9844	20.94%	0.9957
GFS	19.29%	1.0410	19.36%	1.0363	19.92%	1.0056	21.09%	0.9535	21.02%	0.9785	20.83%	0.9948
GFC	19.80%	1.0153	19.85%	1.0126	20.24%	0.9923	21.12%	0.9529	21.01%	0.9792	20.82%	0.9949
GSC	19.94%	0.9742	19.96%	0.9734	20.07%	0.9670	20.36%	0.9567	20.07%	0.9838	19.86%	0.9956
KFS	18.48%	1.0050	18.52%	1.0029	18.88%	0.9873	19.66%	0.9587	19.57%	0.9821	19.44%	0.9953

Lambda	500		1000	
Model	Cor	MSE	Cor	MSE
G	18.63%	0.9968	18.62%	0.9970
K	18.52%	0.9969	18.52%	0.9970
F	19.12%	0.9965	19.12%	0.9968
S	10.06%	0.9971	10.05%	0.9971
C	8.46%	0.9971	8.46%	0.9971
GK	19.89%	0.9968	19.88%	0.9970
GF	20.80%	0.9967	20.79%	0.9969
GS	18.85%	0.9968	18.84%	0.9970
GC	19.67%	0.9968	19.66%	0.9970
KF	19.42%	0.9968	19.42%	0.9970
KS	18.53%	0.9969	18.53%	0.9971
KC	19.52%	0.9969	19.52%	0.9971
FS	19.19%	0.9966	19.19%	0.9969
FC	19.12%	0.9965	19.12%	0.9968
SC	12.94%	0.9971	12.94%	0.9971
GKF	20.80%	0.9967	20.80%	0.9969
GKS	19.92%	0.9968	19.92%	0.9970
GKC	20.92%	0.9969	20.92%	0.9970
GFS	20.80%	0.9967	20.80%	0.9969
GFC	20.80%	0.9967	20.79%	0.9970
GSC	19.84%	0.9969	19.83%	0.9970
KFS	19.42%	0.9968	19.42%	0.9970

Lambda	0.0001		0.001		0.01		1		10		100	
Model	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE
KFC	19.81%	0.9610	19.82%	0.9608	19.85%	0.9591	19.88%	0.9613	19.65%	0.9866	19.55%	0.9959
KSC	19.86%	0.9606	19.87%	0.9604	19.89%	0.9588	19.90%	0.9614	19.64%	0.9867	19.54%	0.9959
FSC	-0.98%	1355.4548	3.27%	17.9545	13.09%	1.3633	18.92%	0.9778	19.27%	0.9745	19.16%	0.9940
GKFS	19.31%	1.0390	19.38%	1.0345	19.93%	1.0045	21.09%	0.9534	21.02%	0.9786	20.83%	0.9948
GKFC	21.33%	0.9558	21.33%	0.9555	21.38%	0.9532	21.47%	0.9553	21.18%	0.9849	21.04%	0.9957
GKSC	21.21%	0.9578	21.21%	0.9575	21.28%	0.9546	21.39%	0.9549	21.10%	0.9846	20.95%	0.9957
GFSC	20.07%	1.0060	20.11%	1.0037	20.43%	0.9866	21.17%	0.9526	21.02%	0.9796	20.83%	0.9950
KFSC	19.87%	0.9605	19.88%	0.9603	19.90%	0.9587	19.91%	0.9613	19.66%	0.9867	19.56%	0.9959
GKFSC	21.37%	0.9553	21.38%	0.9550	21.42%	0.9529	21.49%	0.9554	21.19%	0.9850	21.05%	0.9957

Lambda	500		1000	
Model	Cor	MSE	Cor	MSE
KFC	19.54%	0.9969	19.54%	0.9971
KSC	19.53%	0.9969	19.53%	0.9971
FSC	19.14%	0.9965	19.14%	0.9969
GKFS	20.80%	0.9967	20.80%	0.9969
GKFC	21.03%	0.9969	21.03%	0.9970
GKSC	20.93%	0.9969	20.93%	0.9970
GFSC	20.80%	0.9967	20.80%	0.9970
KFSC	19.55%	0.9969	19.55%	0.9971
GKFSC	21.04%	0.9969	21.04%	0.9970

Cor: correlation; MSE: mean squared errors

Table S4.3

Table S4.3. Results of 5-fold KRR prediction analysis using alternative models for hip circumference in GS10K.

Lambda	0.0001		0.001		0.01		1		10		100	
Model	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE
G	14.21%	1.0963	14.28%	1.0910	14.77%	1.0530	15.92%	0.9748	16.12%	0.9813	15.98%	0.9928
K	13.30%	1.0275	13.35%	1.0252	13.75%	1.0080	14.90%	0.9730	15.26%	0.9852	15.25%	0.9934
F	-0.74%	4711.1865	0.61%	178.7725	8.11%	1.5779	14.59%	1.0041	15.32%	0.9773	15.30%	0.9919
S	7.34%	1.0384	7.34%	1.0373	7.36%	1.0276	7.45%	0.9952	7.47%	0.9910	7.43%	0.9941
C	0.61%	11.8724	5.10%	1.1858	6.28%	1.0859	6.30%	1.0105	6.30%	0.9913	6.30%	0.9941
GK	14.97%	1.0385	15.03%	1.0355	15.48%	1.0131	16.65%	0.9679	16.93%	0.9826	16.86%	0.9930
GF	15.19%	1.0988	15.47%	1.0849	16.17%	1.0401	17.45%	0.9683	17.52%	0.9793	17.39%	0.9925
GS	15.20%	1.0380	15.22%	1.0356	15.46%	1.0166	16.14%	0.9709	16.21%	0.9824	16.07%	0.9930
GC	16.46%	1.0016	16.48%	1.0004	16.60%	0.9904	16.97%	0.9662	16.86%	0.9829	16.71%	0.9931
KF	14.05%	1.0492	14.12%	1.0455	14.66%	1.0202	15.73%	0.9724	15.85%	0.9824	15.77%	0.9930
KS	14.09%	1.0063	14.12%	1.0050	14.34%	0.9949	15.04%	0.9722	15.22%	0.9857	15.19%	0.9934
KC	15.98%	0.9778	15.98%	0.9774	16.04%	0.9742	16.16%	0.9691	16.04%	0.9861	15.98%	0.9935
FS	-0.48%	1244.8060	1.44%	13.1344	9.10%	1.5930	14.80%	0.9926	15.35%	0.9787	15.32%	0.9922
FC	-0.74%	4711.1865	0.61%	178.7725	8.11%	1.5779	14.59%	1.0041	15.32%	0.9773	15.30%	0.9919
SC	9.19%	1.0105	9.20%	1.0098	9.28%	1.0035	9.51%	0.9865	9.57%	0.9906	9.55%	0.9941
GKF	5.09%	3.4097	15.51%	1.0815	16.18%	1.0390	17.45%	0.9682	17.52%	0.9793	17.39%	0.9925
GKS	15.69%	1.0142	15.72%	1.0124	15.99%	0.9983	16.77%	0.9670	16.91%	0.9831	16.83%	0.9931
GKC	17.60%	0.9760	17.61%	0.9755	17.68%	0.9710	17.85%	0.9634	17.68%	0.9841	17.58%	0.9933
GFS	15.81%	1.0688	15.88%	1.0635	16.40%	1.0301	17.48%	0.9675	17.52%	0.9795	17.38%	0.9926
GFC	16.35%	1.0401	16.40%	1.0372	16.75%	1.0150	17.56%	0.9661	17.54%	0.9801	17.41%	0.9927
GSC	16.65%	0.9922	16.66%	0.9912	16.76%	0.9835	17.07%	0.9656	16.93%	0.9835	16.79%	0.9932
KFS	14.58%	1.0311	14.63%	1.0287	14.98%	1.0106	15.78%	0.9716	15.83%	0.9828	15.76%	0.9930

Lambda	500		1000	
Model	Cor	MSE	Cor	MSE
G	15.96%	0.9942	15.96%	0.9944
K	15.25%	0.9943	15.25%	0.9944
F	15.30%	0.9940	15.30%	0.9943
S	7.43%	0.9944	7.42%	0.9945
C	6.30%	0.9944	6.30%	0.9945
GK	16.85%	0.9942	16.85%	0.9944
GF	17.37%	0.9941	17.37%	0.9943
GS	16.05%	0.9942	16.05%	0.9944
GC	16.69%	0.9942	16.69%	0.9944
KF	15.76%	0.9942	15.76%	0.9944
KS	15.18%	0.9943	15.18%	0.9944
KC	15.97%	0.9943	15.97%	0.9944
FS	15.31%	0.9941	15.31%	0.9943
FC	15.30%	0.9940	15.30%	0.9943
SC	9.55%	0.9944	9.55%	0.9945
GKF	17.38%	0.9941	17.37%	0.9943
GKS	16.82%	0.9943	16.82%	0.9944
GKC	17.56%	0.9943	17.56%	0.9944
GFS	17.37%	0.9941	17.36%	0.9943
GFC	17.39%	0.9942	17.39%	0.9944
GSC	16.77%	0.9943	16.77%	0.9944
KFS	15.75%	0.9942	15.75%	0.9944

Lambda	0.0001		0.001		0.01		1		10		100	
Model	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE
KFC	16.03%	0.9770	16.03%	0.9766	16.08%	0.9737	16.18%	0.9691	16.05%	0.9862	15.99%	0.9935
KSC	16.03%	0.9767	16.04%	0.9764	16.08%	0.9735	16.17%	0.9692	16.03%	0.9862	15.96%	0.9935
FSC	-1.35%	1409.3999	1.19%	17.5975	9.46%	1.3906	14.66%	1.0005	15.33%	0.9777	15.31%	0.9920
GKFS	15.83%	1.0664	15.90%	1.0615	16.41%	1.0289	17.48%	0.9674	17.52%	0.9796	17.38%	0.9926
GKFC	17.72%	0.9725	17.72%	0.9721	17.77%	0.9685	17.88%	0.9636	17.69%	0.9845	17.60%	0.9933
GKSC	17.64%	0.9749	17.64%	0.9744	17.71%	0.9703	17.85%	0.9635	17.67%	0.9842	17.57%	0.9933
GFSC	16.56%	1.0304	16.60%	1.0279	16.89%	1.0090	17.58%	0.9656	17.54%	0.9804	17.41%	0.9927
KFSC	16.06%	0.9764	16.07%	0.9761	16.11%	0.9733	16.19%	0.9691	16.05%	0.9862	15.99%	0.9935
GKFSC	17.74%	0.9720	17.74%	0.9715	17.79%	0.9682	17.89%	0.9637	17.69%	0.9846	17.59%	0.9933

Lambda	500		1000	
Model	Cor	MSE	Cor	MSE
KFC	15.98%	0.9943	15.98%	0.9944
KSC	15.95%	0.9943	15.95%	0.9944
FSC	15.31%	0.9940	15.30%	0.9943
GKFS	17.37%	0.9941	17.36%	0.9943
GKFC	17.59%	0.9943	17.59%	0.9944
GKSC	17.55%	0.9943	17.55%	0.9944
GFSC	17.39%	0.9942	17.39%	0.9944
KFSC	15.98%	0.9943	15.98%	0.9944
GKFSC	17.58%	0.9943	17.58%	0.9944

Cor: correlation; MSE: mean squared errors

Table S4.4

Table S4.4. Results of 5-fold KRR prediction analysis using alternative models for HDL in GS10K.

Lambda	0.0001		0.001		0.01		1		10		100	
Model	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE
G	19.07%	1.06	19.15%	1.06	19.76%	1.02	21.20%	0.96	21.31%	0.99	20.92%	1.00
K	16.86%	1.01	16.91%	1.01	17.33%	1.00	18.50%	0.97	18.81%	1.00	18.77%	1.01
F	0.04%	5577.76	-1.51%	342.05	10.05%	1.68	17.90%	1.00	18.40%	0.99	18.31%	1.00
S	9.98%	1.03	9.98%	1.03	9.98%	1.02	9.99%	1.00	9.85%	1.00	9.77%	1.01
C	2.29%	11.95	6.54%	1.20	7.11%	1.10	7.05%	1.02	7.06%	1.00	7.06%	1.01
GK	19.92%	1.01	19.99%	1.01	20.49%	0.99	21.77%	0.96	21.90%	0.99	21.71%	1.01
GF	19.93%	1.06	20.10%	1.05	21.08%	1.01	22.47%	0.96	22.33%	0.99	22.05%	1.00
GS	20.16%	1.01	20.20%	1.01	20.53%	0.99	21.44%	0.96	21.38%	0.99	21.02%	1.01
GC	21.48%	0.98	21.50%	0.98	21.65%	0.97	22.14%	0.96	21.91%	0.99	21.56%	1.01
KF	18.00%	1.03	18.07%	1.03	18.60%	1.00	19.46%	0.97	19.33%	0.99	19.20%	1.01
KS	17.65%	1.00	17.68%	1.00	17.92%	0.99	18.65%	0.97	18.78%	1.00	18.72%	1.01
KC	19.75%	0.97	19.75%	0.97	19.77%	0.97	19.77%	0.97	19.55%	1.00	19.46%	1.01
FS	0.91%	1056.96	3.68%	8.95	12.04%	1.35	18.06%	0.99	18.45%	0.99	18.37%	1.00
FC	0.04%	5577.76	-1.51%	342.05	10.05%	1.68	17.90%	1.00	18.40%	0.99	18.31%	1.00
SC	11.83%	1.01	11.84%	1.01	11.90%	1.00	12.04%	0.99	12.00%	1.00	11.96%	1.01
GKF	18.68%	1.09	20.15%	1.05	21.10%	1.01	22.47%	0.96	22.33%	0.99	22.05%	1.00
GKS	20.66%	0.99	20.70%	0.99	21.01%	0.98	21.89%	0.96	21.87%	0.99	21.67%	1.01
GKC	22.68%	0.96	22.68%	0.96	22.73%	0.96	22.82%	0.96	22.47%	0.99	22.28%	1.01
GFS	20.45%	1.04	20.57%	1.03	21.32%	1.00	22.50%	0.96	22.32%	0.99	22.05%	1.00
GFC	21.31%	1.01	21.37%	1.01	21.77%	0.99	22.57%	0.96	22.36%	0.99	22.09%	1.00
GSC	21.72%	0.97	21.73%	0.97	21.87%	0.97	22.25%	0.96	21.98%	0.99	21.65%	1.01
KFS	18.48%	1.01	18.53%	1.01	18.87%	1.00	19.49%	0.97	19.33%	0.99	19.19%	1.01

Lambda	500		1000	
Model	Cor	MSE	Cor	MSE
G	20.85%	1.01	20.84%	1.01
K	18.77%	1.01	18.77%	1.01
F	18.30%	1.01	18.30%	1.01
S	9.76%	1.01	9.76%	1.01
C	7.06%	1.01	7.06%	1.01
GK	21.68%	1.01	21.67%	1.01
GF	22.02%	1.01	22.01%	1.01
GS	20.96%	1.01	20.95%	1.01
GC	21.50%	1.01	21.50%	1.01
KF	19.18%	1.01	19.18%	1.01
KS	18.71%	1.01	18.71%	1.01
KC	19.45%	1.01	19.45%	1.01
FS	18.35%	1.01	18.35%	1.01
FC	18.30%	1.01	18.30%	1.01
SC	11.95%	1.01	11.95%	1.01
GKF	22.02%	1.01	22.01%	1.01
GKS	21.65%	1.01	21.64%	1.01
GKC	22.25%	1.01	22.25%	1.01
GFS	22.01%	1.01	22.01%	1.01
GFC	22.05%	1.01	22.04%	1.01
GSC	21.60%	1.01	21.60%	1.01
KFS	19.18%	1.01	19.17%	1.01

Lambda	0.0001		0.001		0.01		1		10		100	
Model	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE
KFC	19.78%	0.97	19.79%	0.97	19.80%	0.97	19.78%	0.97	19.55%	1.00	19.46%	1.01
KSC	19.81%	0.97	19.81%	0.97	19.82%	0.97	19.79%	0.97	19.55%	1.00	19.45%	1.01
FSC	-1.54%	2283.78	3.69%	13.20	11.80%	1.41	17.95%	0.99	18.42%	0.99	18.33%	1.00
GKFS	20.50%	1.04	20.61%	1.03	21.34%	1.00	22.50%	0.96	22.32%	0.99	22.05%	1.00
GKFC	22.67%	0.96	22.68%	0.96	22.70%	0.96	22.71%	0.96	22.35%	0.99	22.19%	1.01
GKSC	22.72%	0.96	22.72%	0.96	22.76%	0.96	22.82%	0.96	22.46%	0.99	22.27%	1.01
GFSC	21.52%	1.00	21.57%	1.00	21.90%	0.98	22.60%	0.96	22.36%	0.99	22.09%	1.00
KFSC	19.81%	0.97	19.81%	0.97	19.82%	0.97	19.78%	0.97	19.55%	1.00	19.46%	1.01
GKFSC	22.68%	0.96	22.69%	0.96	22.71%	0.96	22.70%	0.96	22.34%	0.99	22.18%	1.01

Lambda	500		1000	
Model	Cor	MSE	Cor	MSE
KFC	19.45%	1.01	19.45%	1.01
KSC	19.44%	1.01	19.44%	1.01
FSC	18.32%	1.01	18.32%	1.01
GKFS	22.01%	1.01	22.01%	1.01
GKFC	22.17%	1.01	22.17%	1.01
GKSC	22.25%	1.01	22.24%	1.01
GFSC	22.05%	1.01	22.05%	1.01
KFSC	19.45%	1.01	19.45%	1.01
GKFSC	22.16%	1.01	22.16%	1.01

Cor: correlation; MSE: mean squared errors

Table S4.5

Table S4.5. Results of 5-fold KRR prediction analysis using alternative models for TC in GS10K.

Lambda	0.0001		0.001		0.01		1		10		100	
Model	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE
G	10.03%	1.1510	10.08%	1.1447	10.47%	1.0999	11.41%	1.0033	11.71%	0.9938	11.53%	1.0015
K	11.57%	1.0501	11.61%	1.0475	11.97%	1.0281	12.84%	0.9880	13.00%	0.9952	12.95%	1.0018
F	2.41%	6546.8922	1.38%	503.0443	6.87%	1.8320	10.71%	1.0372	11.07%	0.9919	11.05%	1.0010
S	9.36%	1.0310	9.36%	1.0300	9.36%	1.0221	9.37%	0.9970	9.24%	0.9983	9.17%	1.0022
C	-0.24%	353.2095	-0.09%	3.8430	0.44%	1.1790	0.53%	1.0473	0.52%	1.0039	0.52%	1.0028
GK	11.65%	1.0695	11.70%	1.0662	12.09%	1.0415	13.04%	0.9891	13.27%	0.9941	13.20%	1.0017
GF	0.15%	361.8796	2.14%	75.0981	-0.14%	120.3638	12.79%	1.0009	12.75%	0.9918	12.67%	1.0012
GS	11.52%	1.0751	11.53%	1.0725	11.67%	1.0513	12.12%	0.9950	12.20%	0.9941	11.99%	1.0016
GC	10.63%	1.0492	10.64%	1.0477	10.74%	1.0346	11.19%	0.9953	11.39%	0.9954	11.26%	1.0018
KF	11.80%	1.0780	11.86%	1.0739	12.22%	1.0460	12.74%	0.9922	12.61%	0.9937	12.52%	1.0016
KS	12.57%	1.0239	12.59%	1.0225	12.75%	1.0115	13.21%	0.9860	13.23%	0.9954	13.16%	1.0019
KC	12.51%	1.0021	12.51%	1.0017	12.53%	0.9976	12.54%	0.9872	12.41%	0.9966	12.35%	1.0020
FS	1.15%	1645.7543	3.03%	12.7490	8.15%	1.4091	11.15%	1.0212	11.36%	0.9923	11.33%	1.0012
FC	2.41%	6546.8922	1.38%	503.0443	6.87%	1.8320	10.71%	1.0372	11.07%	0.9919	11.05%	1.0010
SC	7.96%	1.0245	7.97%	1.0238	7.97%	1.0173	7.99%	0.9984	7.97%	0.9996	7.96%	1.0024
GKF	5.80%	2.5601	11.74%	1.1345	12.10%	1.0840	12.80%	0.9974	12.88%	0.9924	12.78%	1.0013
GKS	12.40%	1.0466	12.42%	1.0446	12.62%	1.0281	13.19%	0.9880	13.28%	0.9942	13.17%	1.0017
GKC	12.50%	1.0119	12.51%	1.0112	12.56%	1.0049	12.73%	0.9873	12.72%	0.9956	12.66%	1.0019
GFS	12.08%	1.1180	12.10%	1.1117	12.36%	1.0724	12.90%	0.9960	12.94%	0.9924	12.83%	1.0014
GFC	11.95%	1.0880	11.97%	1.0845	12.15%	1.0585	12.58%	0.9956	12.65%	0.9930	12.56%	1.0015
GSC	11.32%	1.0331	11.33%	1.0319	11.40%	1.0219	11.71%	0.9920	11.80%	0.9955	11.65%	1.0018
KFS	12.37%	1.0566	12.39%	1.0539	12.59%	1.0343	12.89%	0.9906	12.72%	0.9939	12.62%	1.0016

Lambda	500		1000	
Model	Cor	MSE	Cor	MSE
G	11.48%	1.0025	11.48%	1.0027
K	12.95%	1.0026	12.95%	1.0027
F	11.05%	1.0024	11.05%	1.0026
S	9.16%	1.0027	9.16%	1.0028
C	0.52%	1.0028	0.52%	1.0028
GK	13.19%	1.0026	13.19%	1.0027
GF	12.66%	1.0025	12.66%	1.0026
GS	11.95%	1.0026	11.95%	1.0027
GC	11.24%	1.0026	11.23%	1.0027
KF	12.51%	1.0026	12.50%	1.0027
KS	13.15%	1.0026	13.15%	1.0027
KC	12.34%	1.0027	12.34%	1.0027
FS	11.32%	1.0025	11.32%	1.0026
FC	11.05%	1.0024	11.05%	1.0026
SC	7.95%	1.0027	7.95%	1.0028
GKF	12.76%	1.0025	12.76%	1.0027
GKS	13.15%	1.0026	13.15%	1.0027
GKC	12.65%	1.0026	12.64%	1.0027
GFS	12.81%	1.0025	12.81%	1.0027
GFC	12.54%	1.0025	12.54%	1.0027
GSC	11.62%	1.0026	11.62%	1.0027
KFS	12.61%	1.0026	12.61%	1.0027

Lambda	0.0001		0.001		0.01		1		10		100	
Model	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE	Cor	MSE
KFC	12.44%	1.0018	12.44%	1.0013	12.45%	0.9975	12.46%	0.9874	12.33%	0.9967	12.28%	1.0020
KSC	12.70%	1.0001	12.70%	0.9996	12.70%	0.9960	12.69%	0.9867	12.53%	0.9966	12.47%	1.0020
FSC	1.56%	8428.8690	3.02%	18.9609	7.92%	1.5159	10.84%	1.0324	11.14%	0.9920	11.12%	1.0010
GKFS	12.09%	1.1150	12.11%	1.1091	12.38%	1.0710	12.92%	0.9957	12.95%	0.9925	12.85%	1.0014
GKFC	12.64%	1.0070	12.64%	1.0064	12.67%	1.0011	12.79%	0.9868	12.77%	0.9959	12.72%	1.0019
GKSC	12.66%	1.0097	12.67%	1.0090	12.70%	1.0031	12.84%	0.9868	12.81%	0.9956	12.74%	1.0019
GFSC	12.21%	1.0752	12.23%	1.0723	12.35%	1.0502	12.69%	0.9942	12.73%	0.9931	12.63%	1.0015
KFSC	12.47%	1.0011	12.47%	1.0006	12.48%	0.9969	12.48%	0.9873	12.35%	0.9967	12.29%	1.0021
GKFSC	12.70%	1.0060	12.70%	1.0054	12.73%	1.0004	12.83%	0.9866	12.80%	0.9959	12.75%	1.0019

Lambda	500		1000	
Model	Cor	MSE	Cor	MSE
KFC	12.27%	1.0027	12.27%	1.0027
KSC	12.46%	1.0027	12.46%	1.0027
FSC	11.12%	1.0024	11.12%	1.0026
GKFS	12.83%	1.0025	12.83%	1.0027
GKFC	12.71%	1.0026	12.71%	1.0027
GKSC	12.73%	1.0026	12.73%	1.0027
GFSC	12.61%	1.0025	12.61%	1.0027
KFSC	12.28%	1.0027	12.28%	1.0027
GKFSC	12.75%	1.0026	12.74%	1.0027

Cor: correlation; MSE: mean squared errors

Table S5.1

Table S5.1 Results of sib-sib and sib-in-law pedigree study in GS20K

Trait	r_{FS}			r_{FSIL}			r_{CP}		
	Estimate	s.e.	d.f.	Estimate	s.e.	d.f.	Estimate	s.e.	d.f.
Creatinine	0.3491	0.0222	1,779	0.0869	0.0236	1,776	0.1452	0.0233	1,798
Hip	0.2834	0.0224	1,828	0.0857	0.0233	1,836	0.1157	0.0232	1,840
DBP	0.1624	0.0228	1,866	0.0880	0.0230	1,873	0.1640	0.0228	1,873
HR	0.1388	0.0230	1,859	0.0601	0.0231	1,864	0.1245	0.0230	1,863
Height	0.5202	0.0197	1,875	0.1709	0.0227	1,880	0.2731	0.0222	1,882
Fat	0.3034	0.0224	1,806	0.0502	0.0235	1,799	0.2013	0.0230	1,809
Weight	0.3130	0.0220	1,870	0.0661	0.0231	1,873	0.1873	0.0227	1,876
Waist	0.2421	0.0226	1,838	0.0696	0.0232	1,843	0.1924	0.0228	1,846
BMI	0.3120	0.0220	1,867	0.0917	0.0230	1,871	0.2358	0.0224	1,875
Urea	0.1970	0.0232	1,783	-0.0140	0.0237	1,780	0.0851	0.0235	1,803
TC	0.1615	0.0234	1,780	0.0087	0.0237	1,775	0.0485	0.0235	1,806
HDL	0.2929	0.0227	1,777	0.0449	0.0237	1,772	0.1070	0.0234	1,798
Glucose	0.1725	0.0239	1,705	0.0287	0.0243	1,693	0.1314	0.0240	1,711
WHR	0.1389	0.0232	1,825	0.0286	0.0233	1,833	0.1265	0.0231	1,838
SBP	0.1522	0.0229	1,865	0.0287	0.0231	1,872	0.0612	0.0231	1,871

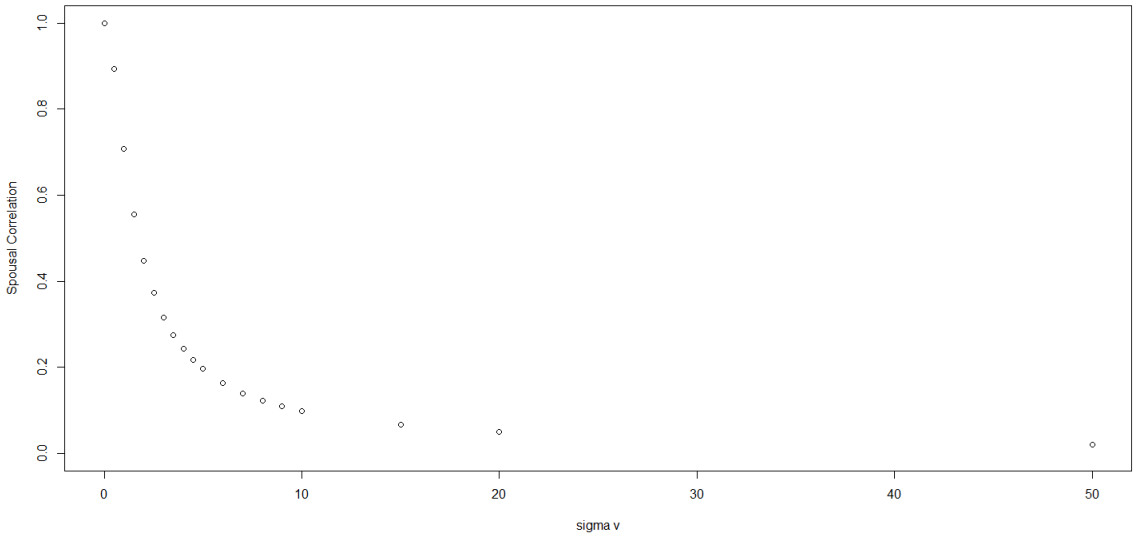
Trait	ρ		h^2_{AM}		h^2_{gkin}	
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
Creatinine	0.2489	0.0676	0.6066	0.0951	0.6000	0.0300
Hip	0.3024	0.0822	0.4933	0.0771	0.4200	0.0400
DBP	0.5419	0.1416	0.2817	0.0425	0.1600	0.0200
HR	0.4330	0.1664	0.2504	0.0534	0.2600	0.0300
Height	0.3285	0.0436	0.8197	0.0696	0.8800	0.0300
Fat	0.1655	0.0775	0.5558	0.1423	0.4900	0.0400
Weight	0.2112	0.0738	0.5599	0.1094	0.5400	0.0400
Waist	0.2875	0.0958	0.4309	0.0809	0.5000	0.0300
BMI	0.2939	0.0737	0.5387	0.0785	0.4800	0.0400
Urea	-0.0711	0.1203	0.4057	0.3335	0.2500	0.0300
TC	0.0539	0.1467	0.3176	0.4403	0.3300	0.0300
HDL	0.1533	0.0809	0.5410	0.1548	0.5600	0.0300
Glucose	0.1664	0.1409	0.3272	0.1461	0.1700	0.0200
WHR	0.2059	0.1677	0.2635	0.1133	0.3000	0.0300
SBP	0.1886	0.1518	0.2886	0.1225	0.1900	0.0300

r_{FS}, r_{FSIL}, r_{CP} refer to observed phenotypic correlation between full-siblings, between sibs-and sibs-in-law and between partners respectively. Some sib pairs and couple pairs are used more than once when estimating r_{FS} and r_{CP} respectively, if they have more than one sibs-in-law in the data.

SP Figures

Figure S5.1

Figure S5.1. Empirical distribution of Sigma (X-axis) and Spousal phenotypic correlation (Y-axis).



Assortatively mate 50K simulated males and females using the mating process described in Chapter 5 but with different sigma ranging from 0 to 50. Each dot is the mean of 50 replications. $VarP$ is always 1 here.

Figure S5.2

Figure S5.2. Observed and Expected Familial Resemblances between Different Types of Relatives within Nuclear Family under Assortative Mating

●, ●, ●, ●, ●, ●, ● and ● refer to simulated assortative mating populations in which

$VarG_0$	$\rho_0 = 0.15$	$\rho_0 = 0.30$
0.20	●	●
0.40	●	●
0.60	●	●
0.65	●	●

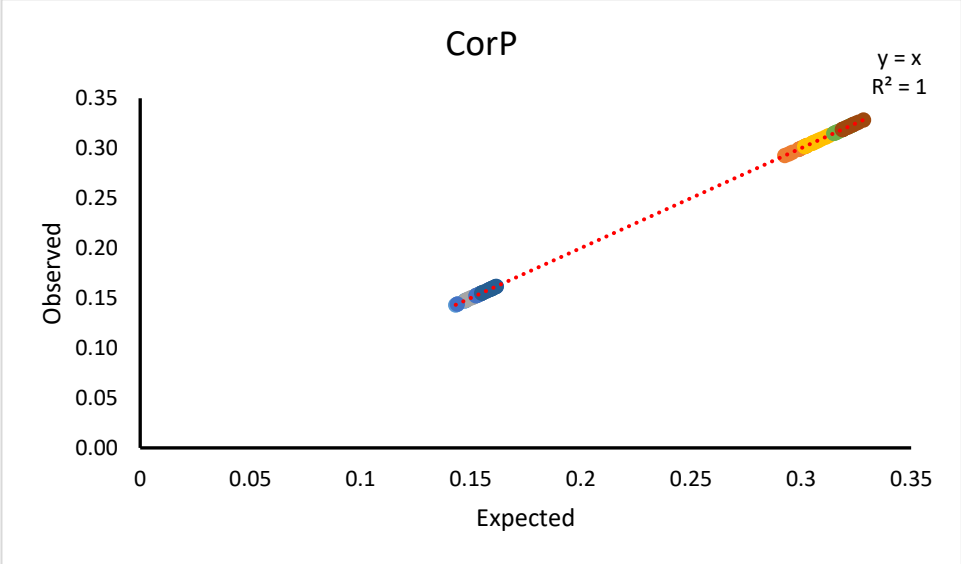
If the expectation is a positive value, I regressed the observation on the expectation to see whether the regression coefficient and R^2 are close to 1.

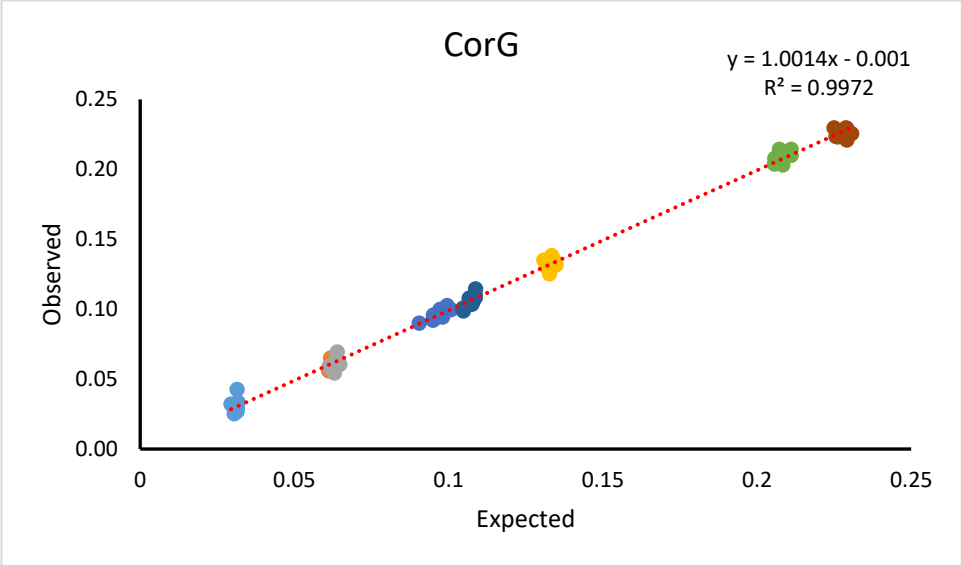
If the expectation is 0, I conducted sign test to see whether there is any bias in the sign of direction, e.g. more positive values than negative.

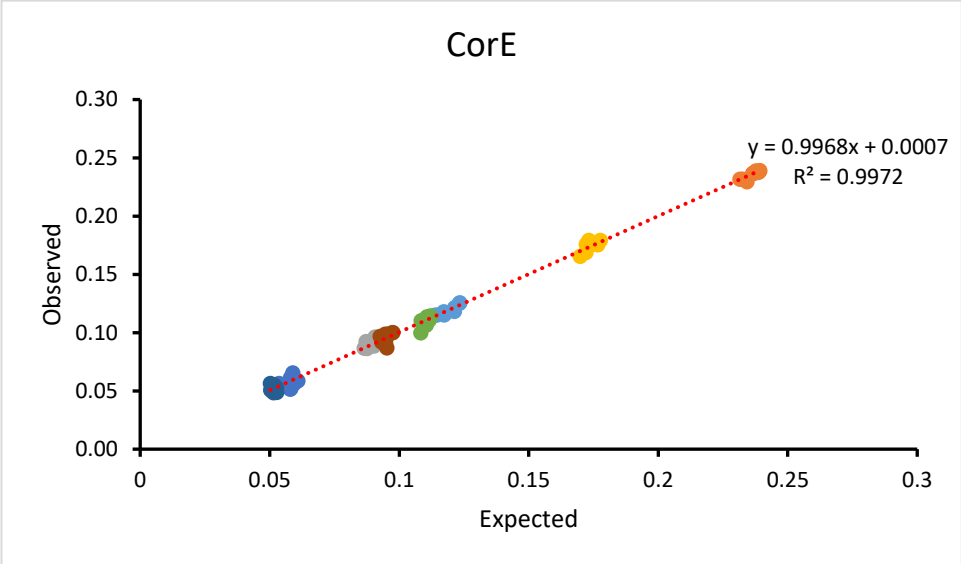
Expectations of familial resemblance between different types of 1st degree relatives see Table 5.1 in Chapter 5.

CorP, CorG and CorE refer to the correlations between phenotypic values, between genetic values and between environmental values respectively.

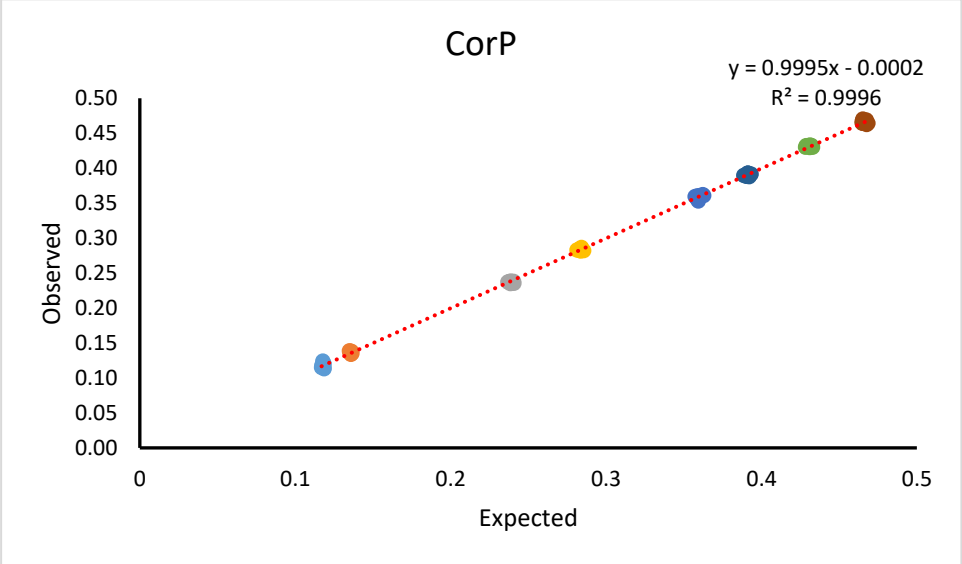
Observed vs Expected familial resemblance between partners

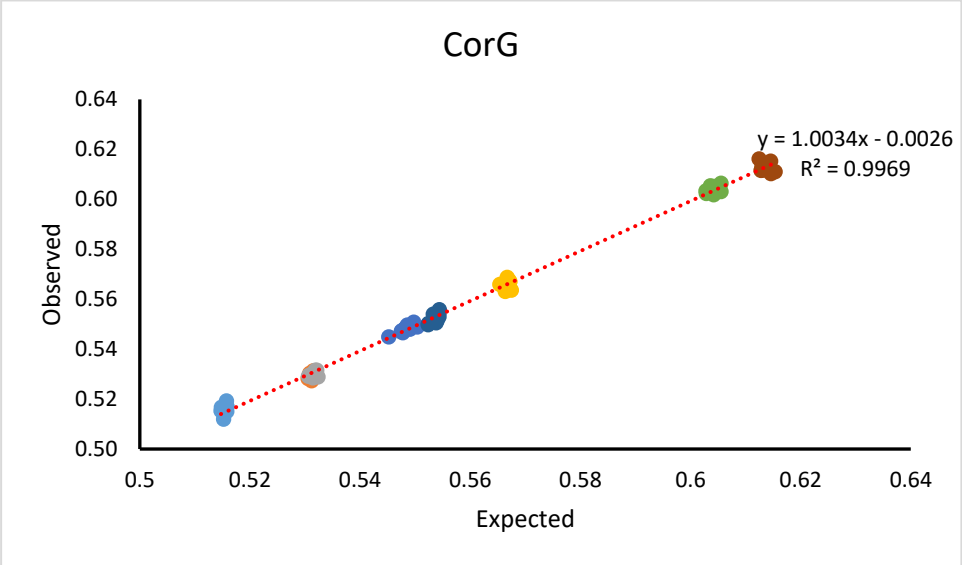


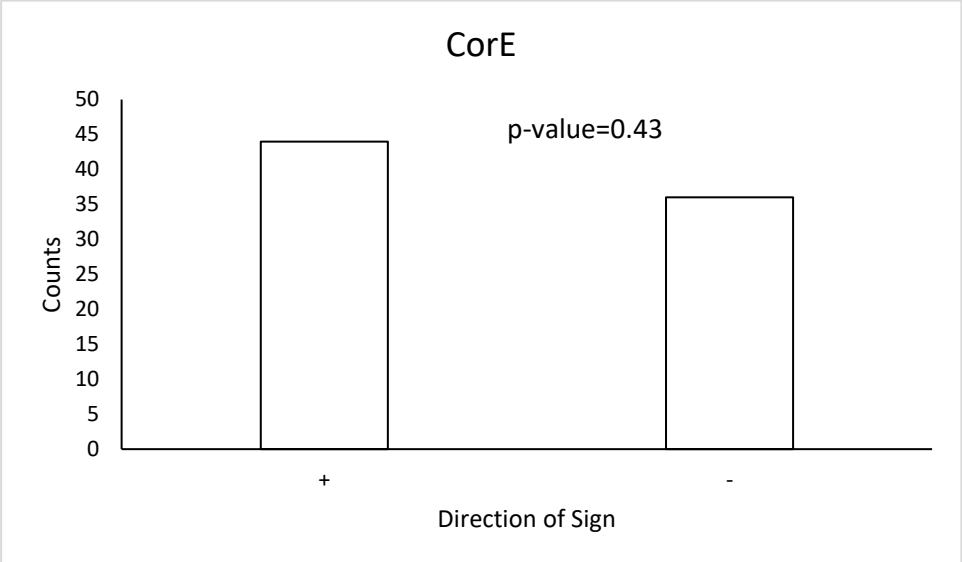




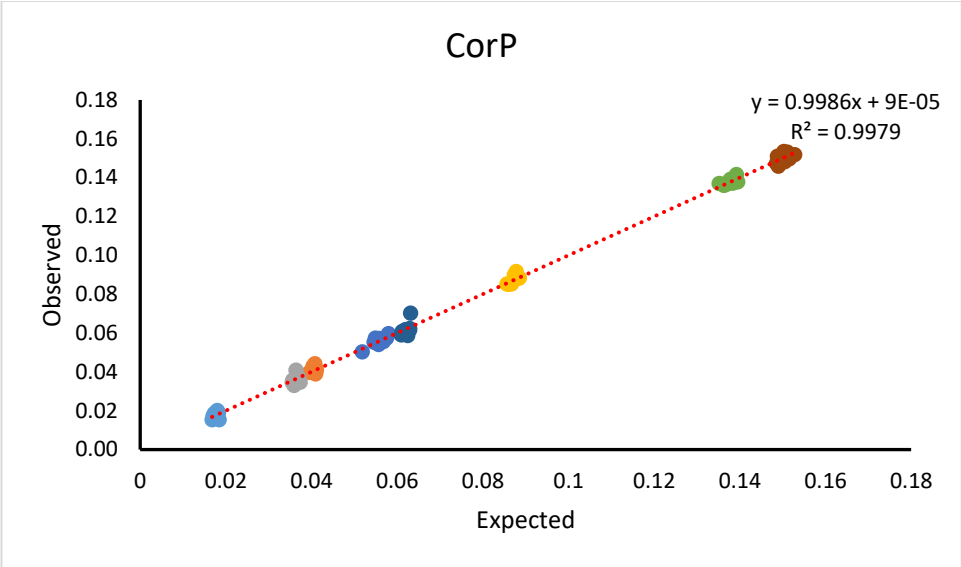
Observed vs Expected familial resemblance between parents and offspring

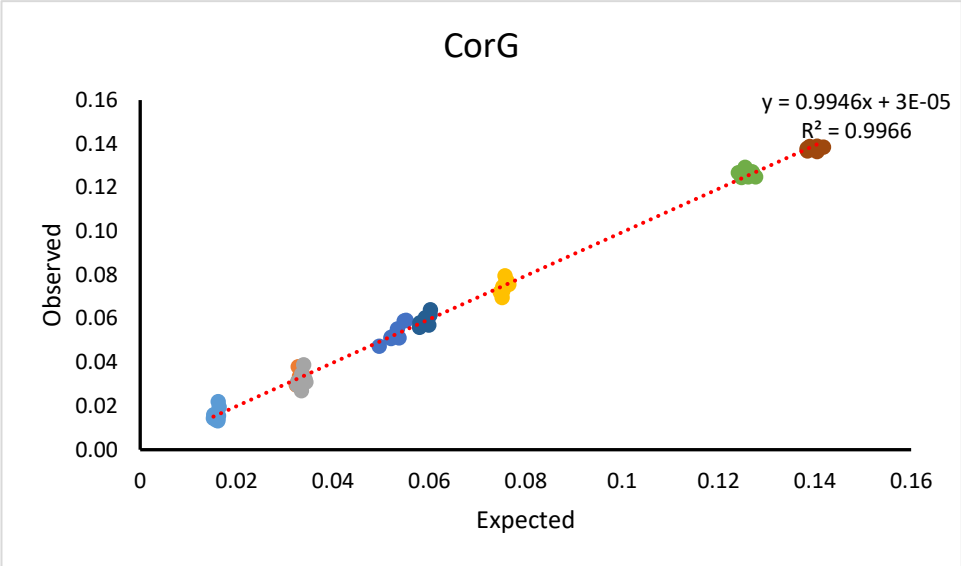


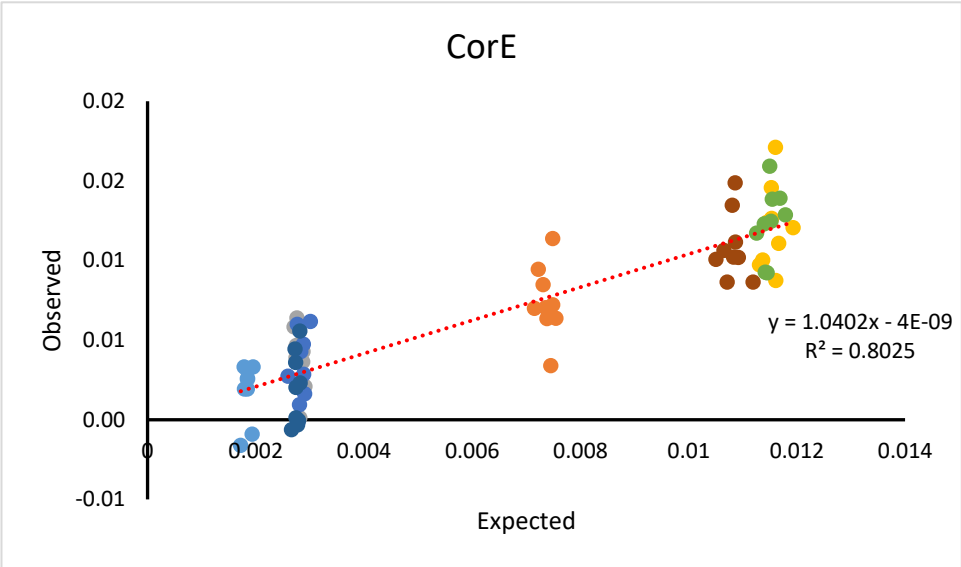




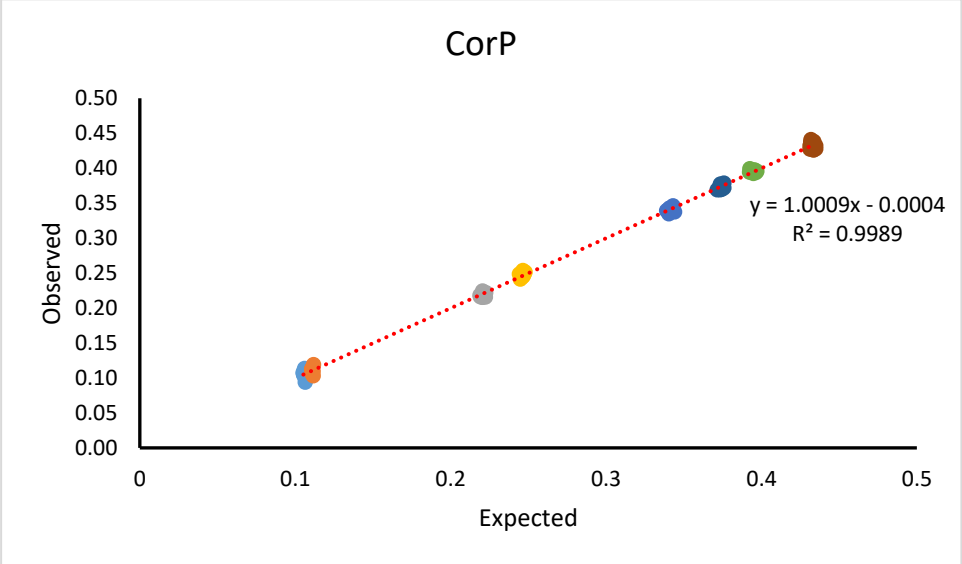
Observed vs Expected familial resemblance between parents and offspring-in-law

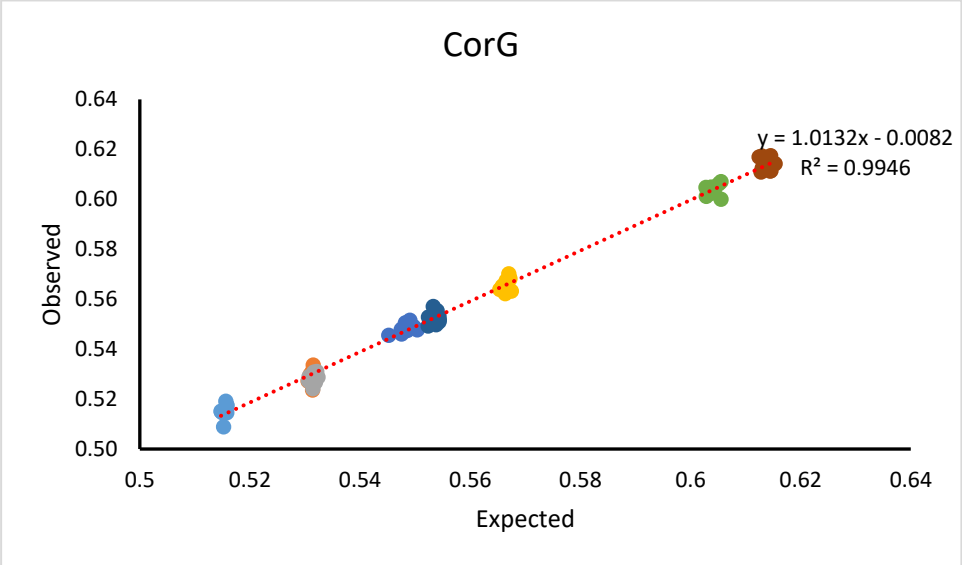


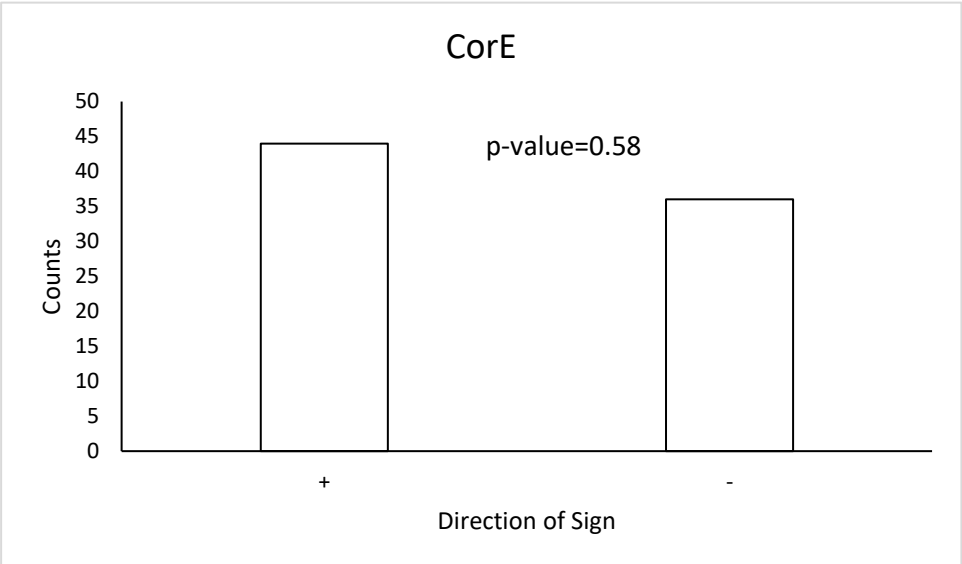




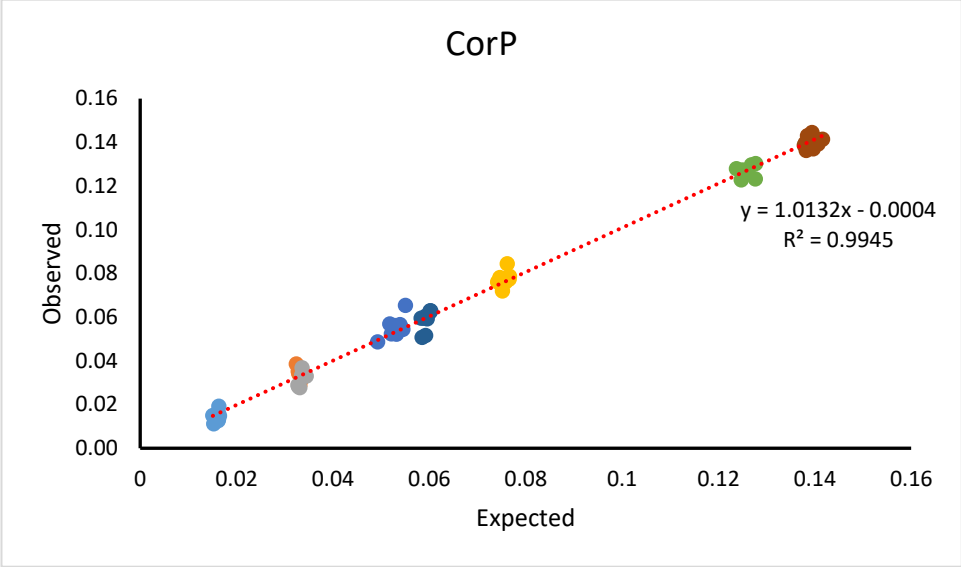
Observed vs Expected familial resemblance between full-siblings

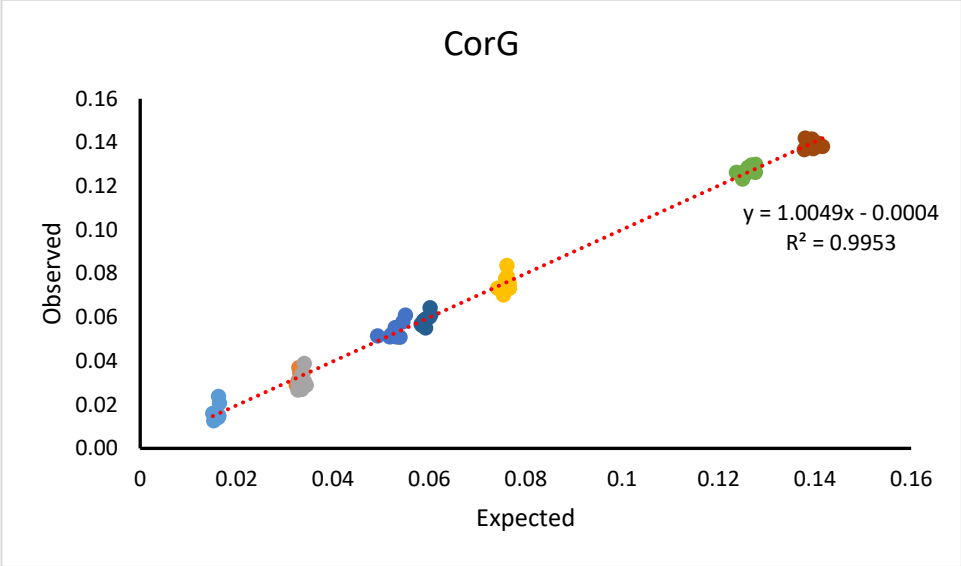






Observed vs Expected familial resemblance between full-siblings and siblings-in-law





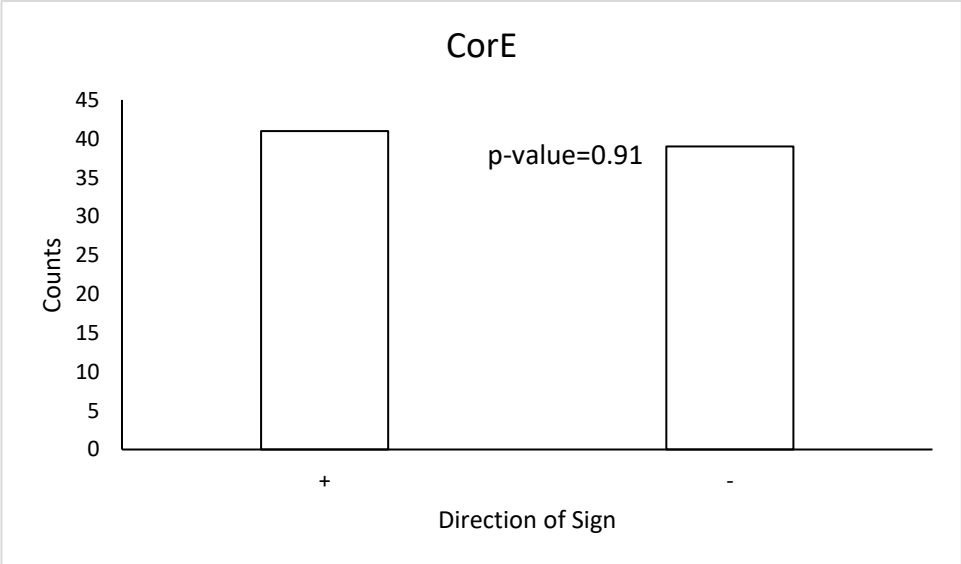
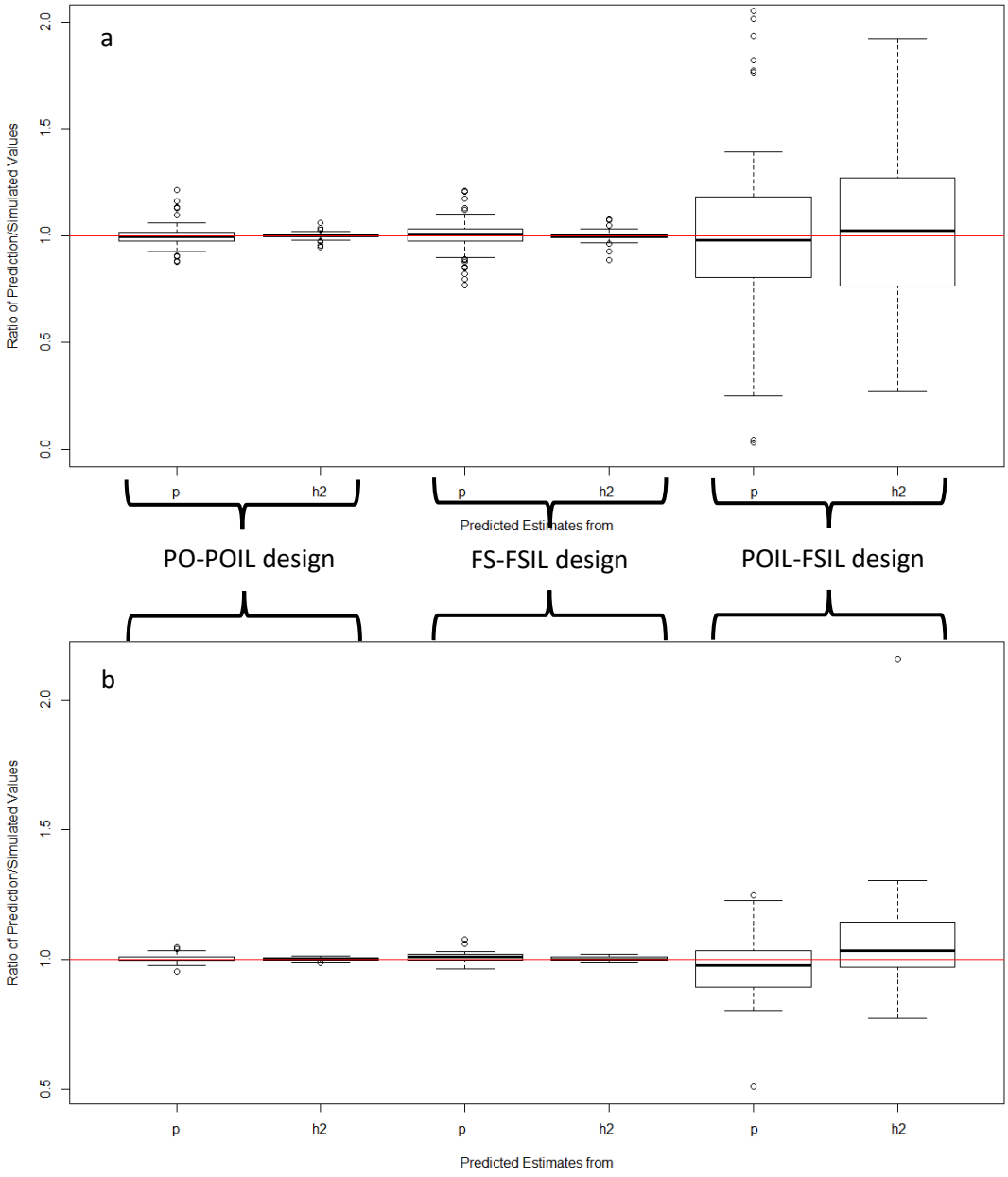


Figure S5.3

Figure S5.3 Comparing the predicted intensity of assortative mating and heritability to their simulated values.



For each generation in the last 10 generations of simulated assortative mating cohorts (equilibrium reached), I predicted ρ and h^2_{AM} using parent-offspring and POIL relationships, full-sibling and FSIL relationships and POIL and FSIL relationships, abbreviated as PO-POIL, FS-FSIL and POIL-FSIL design accordingly, based on equations 25 to 30 in Chapter 5. And then I estimated the ratio of the predicted values and simulated values (the observed couple correlation in the previous generation and the observed heritability in current generation) and boxplot the ratios in figure a. Red dash line: y=1. On average, the predicted values are unbiased (medians overlap with

red dash line). However, although the overall predicted values are unbiased, the standard errors of ρ and h^2_{AM} predicted from POIL-FSIL design are very large. This figure has been zoomed in, the origin range of ratio of prediction/simulated value is -0.28 to 3.31 for ρ and -9.97 to 33.47 for h^2_{AM} in POIL-FSIL design. Significant deviations in POIL-FSIL design are mainly caused by traits with low founder heritability and low intensity of assortative mating (e.g. $VarG_0 = 0.2$ and $\rho_0 = 0.15$). Therefore, I made a same plot (figure b) but only kept three cohorts in which the intensity of assortative mating is 0.3 and the founder heritability is ≥ 0.4. The performance of POIL-FSIL design is better in plot b as the deviations are smaller.

SP Publications

Publication S2.1

Shared genetics and couple-associated environment are major contributors to the risk of both clinical and self-declared depression

Yanni Zeng (MSc)[1*], Pau Navarro (PhD)[2], Charley Xia (MSc)[2], Carmen Amador (PhD)[2], Ana M. Fernandez-Pujals (MSc)[1], Pippa A. Thomson (PhD)[4,5], Archie Campbell (MA)[3], Reka Nagy (MSc)[2], Toni-Kim Clarke (PhD)[1], Jonathan D. Hafferty (MD)[1], Blair H. Smith (MD)[3,6], Lynne J. Hocking (PhD)[3,7], Sandosh Padmanabhan (PhD)[3,8], Caroline Hayward (PhD)[2], Donald J. MacIntyre (MD)[1], David J Porteous (PhD)[3,4,5], Chris S. Haley (PhD)[2,9] and Andrew M. McIntosh (MD)[1,3,4]

Affiliations:

[1] Division of Psychiatry, University of Edinburgh, Edinburgh, UK

[2] MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

[3] Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

[4] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, United Kingdom

[5] Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh

[6] Division of Population Health Sciences, University of Dundee, Dundee, UK

[7] Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

[8] Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

[9] The Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, UK

* Corresponding author

For manuscript see:

http://www.sciencedirect.com/science/article/pii/S2352396416305072?via%3Dihub

or scan QR code

https://uoe-my.sharepoint.com/personal/s1231856_ed_ac_uk/Documents/Charley%27s%20Appendix?csf=1&e=r6IA4S

Publication S2.2

Genomic analysis of family data reveals additional genetic effects on intelligence and personality

W. David Hill[1,2]*†, Ruben C. Arslan[3,4]†, Charley Xia†[5], Michelle Luciano[1,2], Carmen Amador[5], Pau Navarro[5], Caroline Hayward[5], Reka Nagy[5], David J. Porteous[1,8,9], Andrew M. McIntosh[1,10], Ian J. Deary[1,2], Chris S. Haley[5,11], and Lars Penke[1,3,4]

[1] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK

[2] Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK

[3] Georg Elias Müller Institute of Psychology, Georg August University Göttingen, Germany

[4] Leibniz ScienceCampus Primate Cognition, Göttingen, Germany

[5] MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

[7] Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

[8] Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

[9] Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh

[10] Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh EH10 5HF

[11] The Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, UK

* Corresponding author

† These authors contributed equally

For manuscript see

https://www.nature.com/articles/s41380-017-0005-1

or

https://uoe-my.sharepoint.com/personal/s1231856_ed_ac_uk/Documents/Charley%27s%20Appendix?csf=1&e=r6IA4S

or scan QR code

Publication S2.3

Regional variation in health is predominantly driven by lifestyle rather than genetics

Carmen Amador[1], Charley Xia[1], Réka Nagy[1], Archie Campbell[2,3], David Porteous[2,3], Blair H. Smith[3,4], Nick Hastie[1], Veronique Vitart[1], Caroline Hayward[1], Pau Navarro[1]*, Chris S. Haley[1,5]*

[1] MRC Human Genetics Unit, Institute of Genetic and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, United Kingdom

[2] Centre for Genomic and Experimental Medicine, Institute of Genetic and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, United Kingdom

[3] Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetic and Experimental Medicine, University of Edinburgh, Edinburgh, EH4 2XU, United Kingdom

[4] Division of Population Health Sciences, University of Dundee, Dundee, DD2 4RB, United Kingdom

[5] Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, EH25 9RG, United Kingdom

* These authors contributed equally to this work

For manuscript see:

https://www.nature.com/articles/s41467-017-00497-5

or

https://uoe-my.sharepoint.com/personal/s1231856_ed_ac_uk/Documents/Charley%27s%20Appendix?csf=1&e=r6IA4S

or scan QR code:

Publication S3.1

New tools for genome-wide association studies on a population level scale

Oriol Canela-Xandri*[,1], Charley Xia*[,2], Konrad Rawlik[1], Carmen Amador[2], Yanni Zeng[2], David Porteous[3], Caroline Hayward[2], Pau Navarro[2], Chris Haley[1,2], Albert Tenesa[1,2,†].

Affiliations:

[1]The Roslin Institute, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG. Scotland. UK.

[2]MRC HGU at the MRC IGMM, University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh. EH4 2XU. UK

[3]Centre for Genomic and Experimental Medicine, Institute of Genetic and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, United Kingdom

*These authors equally contributed to this work.

†Corresponding author

For manuscript see:

https://uoe-my.sharepoint.com/personal/s1231856_ed_ac_uk/Documents/Charley%27s%20Appendix?csf=1&e=r6IA4S

or scan QR code: