



Review in Advance first posted online
on August 22, 2013. (Changes may
still occur before final publication
online and in print.)

Estimation and Partition of Heritability in Human Populations Using Whole-Genome Analysis Methods

Anna A.E. Vinkhuyzen,¹ Naomi R. Wray,¹
Jian Yang,^{1,2} Michael E. Goddard,^{3,4}
and Peter M. Visscher^{1,2}

¹The University of Queensland, Queensland Brain Institute, Brisbane, 4072, Queensland, Australia; email: anna.vinkhuyzen@uq.edu.au, naomi.wray@uq.edu.au, peter.visscher@uq.edu.au

²The University of Queensland, Diamantina Institute, Translation Research Institute, Brisbane, 4072, Queensland, Australia; email: jian.yang@uq.edu.au

³Department of Food and Agricultural Systems, University of Melbourne, Parkville, 3053, Victoria, Australia

⁴Biosciences Research Division, Department of Primary Industries, Bundoora, 3001, Victoria, Australia; email: mike.goddard@dpi.vic.gov.au

Annu. Rev. Genet. 2013. 47:93–113

The *Annual Review of Genetics* is online at
genet.annualreviews.org

This article's doi:
10.1146/annurev-genet-111212-133258

Copyright © 2013 by Annual Reviews.
All rights reserved

Keywords

quantitative traits, whole-genome methods, additive genetic variance, genomic relationship, mixed linear model, genetic architecture

Abstract

Understanding genetic variation of complex traits in human populations has moved from the quantification of the resemblance between close relatives to the dissection of genetic variation into the contributions of individual genomic loci. However, major questions remain unanswered: How much phenotypic variation is genetic; how much of the genetic variation is additive and can be explained by fitting all genetic variants simultaneously in one model; and what is the joint distribution of effect size and allele frequency at causal variants? We review and compare three whole-genome analysis methods that use mixed linear models (MLMs) to estimate genetic variation. In all methods, genetic variation is estimated from the relationship between close or distant relatives on the basis of pedigree information and/or single nucleotide polymorphisms (SNPs). We discuss theory, estimation procedures, bias, and precision of each method and review recent advances in the dissection of genetic variation of complex traits in human populations. Using genome-wide data, it is now established that SNPs in total account for far more of the genetic variation than the statistically highly significant SNPs that have been detected in genome-wide association studies. All SNPs together, however, do not account for all of the genetic variance estimated by pedigree-based methods. We explain possible reasons for this remaining “missing heritability.”

Complex trait: a trait for which variation is a result of multiple genes and environmental factors (quantitative and categorical)

Genetic relationship: the expected or actual proportion of the genome identical by descent (pedigree or within-family design) or by state (population design)

Genetic architecture: description of the number of genetic variants, allelic frequencies, and effect sizes of variants that affect a trait and their mode of gene action

INTRODUCTION

The discipline of quantitative genetics, or the genetics of complex traits, aims to understand and exploit genetic variation in continuously varying traits, such as height (stature), blood pressure, and cognitive ability in humans. However, relatives often share both genes and a similar environment, making it difficult to completely separate the genetic variance from the variance due to the shared environment.

Technological advances now allow individuals to be assayed for more than one million genetic markers [usually single nucleotide polymorphisms (SNPs)] covering DNA variation that spans the whole genome. These SNP data can be used in two ways: to map genes that affect a complex trait and to estimate the genetic relationship between individuals more accurately than can be done from their known pedigree. Estimating genetic relationships between individuals from SNP data allows us to estimate genetic variance from supposedly unrelated individuals without confounding by shared envi-

ronment and allows dissection of the genetic architecture of complex traits. Genetic architecture refers to the description of the number, frequencies, and effect sizes as well as the mode of gene action of causal variants.

Our focus here is on the theory and application of whole-genome analysis methods to estimate genetic variance in human populations and to elucidate the genetic architecture of complex traits. In all of the analysis methods that we review, mixed linear models (MLMs) form the basis of the analysis. The focus is on human populations because large genetic data sets are available, allowing accurate empirical validation of new genomic analysis methods. In the discussion of the analysis methods, we focus on analyses of quantitative traits; variations of these methods have been developed to allow application to discrete traits (see sidebar, Estimating Heritability for Disease).

We start with a concise history to place quantitative genetics and whole-genome methods in context. We then review and compare three different designs and methods for estimation of genetic variance and discuss precision and potential sources of bias of the estimates. The three designs are referred to as the pedigree design, the within-family design, and the population design. Briefly, to estimate heritability, the pedigree design utilizes observed and expected similarity of identical [monozygotic (MZ)] and nonidentical [dizygotic (DZ)] twin pairs, the within-family design utilizes realized variation around expected genetic similarity for full-sibling pairs, and the population design utilizes realized genetic similarity between distant relatives. There are other pedigree designs that can be used to estimate genetic variance, e.g., parents and offspring or extended family pedigrees, but in this review we focus on the twin design. Similarly, there are other within-family designs that could be used, e.g., families with half-siblings, but here we focus on families with full siblings. For each method, we summarize an example study on the model trait human height. We chose to select examples for human height because height is a highly (~80%) heritable trait (e.g., see 70, 85),

ESTIMATING HERITABILITY FOR DISEASE

Categorical traits (e.g., disease traits) are sometimes inherited in a simple Mendelian manner, but often they behave like continuous traits in that they are influenced by many genes and environmental factors. Unlike quantitative traits, the phenotypic variance of a disease trait depends on the population mean, i.e., the disease prevalence. To facilitate comparison of estimates across studies, heritability of disease is generally estimated for the liability underlying the disease (48) in a threshold model (21). In a threshold model, the disease status (measured on a 0/1 scale) is superimposed onto a distribution of liability in which a threshold of liability for disease is determined that bisects the distribution to reflect the proportion of affected and unaffected individuals in the population (20).

Estimates derived from linear mixed model analysis can be transformed to a liability scale by adjusting both for scale and for ascertainment of the data. Estimation methods for heritability of disease have recently been reviewed by Tenesa & Haley (75), and a detailed description of estimating heritability for disease traits in a population design is provided by Lee et al. (45).

has been studied for more than century (24), and large empirical data sets are available to demonstrate analysis methods and statistical inference. Finally, we discuss to what extent the whole-genome methods have contributed to a better understanding of the genetic architecture of complex traits in human populations and how developments in analysis methods and DNA sequencing technology can contribute further.

A BRIEF HISTORY OF COMPLEX TRAIT GENETICS IN HUMAN POPULATIONS

Quantitative genetic methods build on the principles of genetics described by Mendel and on statistical methods developed initially by Galton. Whereas Mendel in his experiments with peas focused on discrete traits, Galton pioneered statistical methods to study the resemblance between relatives and introduced concepts of regression and correlation to study continuous variation in a population. In his book *Hereditary Genius* (23), Galton stated that if a trait is heritable, then the closer the familial relatedness of two individuals, the more these people phenotypically resemble each other. Several articles have been published on continuous distribution of quantitative traits and the observed resemblance between relatives (7, 59, 96), and a comprehensive theory to explain these observations was presented in R.A. Fisher's landmark paper "The Correlation Between Relatives on the Supposition of Mendelian Inheritance," published in 1918 (22). Fisher showed that the seemingly contradictory hereditary properties of discrete traits (e.g., Mendel's peas) and continuous traits (e.g., Galton's height) were consistent if quantitative trait variation is caused by a combination of many genetic loci, each with a small effect and inherited in a Mendelian manner, together with environmental effects. The article by Fisher marks the beginning of the discipline of quantitative genetics. According to Fisher's infinitesimal model, many segregating genes, each with a very small (infinitesimal) effect on the trait, lead to a normal distribution of genetic

values and, provided environmental effects are normally distributed, a normal distribution of phenotypes in the population. The theory implied that genetic and nongenetic sources of variation can be estimated by quantifying the correlation between relatives, without any knowledge of specific genes affecting the trait. Further important theoretical developments on the genetics, selection, and evolution of quantitative traits were made by Wright (90), Crow & Kimura (10), and many others (e.g., 21, 49).

With the advent of molecular genetics, quantitative genetics, as with many other disciplines in biology, has become a more empirical, data-driven science. New data can be used to answer old questions about the genetic architecture of complex traits. Whereas, for example, in the past expected values of genetic similarity were used to estimate heritability from relatives, it is now possible to estimate empirically the realized genetic similarity between close or distant relatives and to estimate heritability exploiting that information.

SOURCES OF VARIATION

Total phenotypic variation between individuals (V_P) is a directly observable measure in a population. Many systematic and random factors are expected to contribute to that variation. We can partition total phenotypic variation into a genetic (V_G) and non-genetic (V_R) component of variation:

$$V_P = V_G + V_R. \quad (1)$$

Equation 1 represents the simplest partitioning of V_P in which no specific factors are identified that contribute to V_G and V_R . For a complex trait, however, many genetic and environmental factors are likely to contribute to the variance, and V_G and V_R can be partitioned further.

We may partition V_R into variance due to effects shared by children in a nuclear family that live together (common environment, V_C) and a remaining residual variance (V_E), although further variance components could be partitioned out from the residual variance, for example, to

Whole-genome analysis methods:

methods that utilize information from the whole genome to estimate genetic parameters, often in combination with phenotypic information

Mixed linear model (MLM):

a linear model that jointly accounts for fixed and random effects



account for genotype by environment interaction and measurement error:

$$V_R = V_C + V_E. \quad (2)$$

Similarly, the genetic component for a complex trait likely comprises multiple genetic factors. These genetic factors may simply add up but could also interact with each other (i.e., non-additivity). Nonadditivity refers to interaction between alleles at the same locus (dominance) or at different loci (epistasis). Under this model, the total genetic variance of a complex trait is defined as the sum of all additive and nonadditive genetic components:

$$V_G = V_A + V_D + V_I, \quad (3)$$

where V_G is the total genetic variance, V_A is the additive genetic variance, V_D is the dominance variance, and V_I is the epistatic variance. V_D and V_I are collectively referred to as non-additive genetic variance (21, 49). Nonadditive variance can be further partitioned in interactions between additive and nonadditive variance components and between nonadditive and non-additive variance components:

$$V_G = V_A + V_D + V_{AA} + V_{AD} + V_{AAA} + V_{DD} + V_{AAD} + \dots \quad (4)$$

We can consider the covariance between pairs of relatives. For example, the expected covariance between MZ twins and DZ twins (or full siblings) is respectively (21, 49):

$$Cov_{mz} = V_G + V_C \quad (5)$$

and

$$Cov_{dz} = \frac{1}{2}V_A + \frac{1}{4}V_D + \frac{1}{4}V_{AA} + \frac{1}{8}V_{AD} + \frac{1}{8}V_{AAA} + \frac{1}{16}V_{DD} + \frac{1}{16}V_{AAD} + \dots + V_C. \quad (6)$$

Therefore, theoretically, many variance components can be partitioned from the expected covariance between close relatives. However, because there are in theory more components than data points and because a number of these components are confounded (e.g., V_D and V_{AA} in the covariance between DZ twin pairs or full

sibling pairs), not all components are estimable with current experimental designs.

In this review, we focus on estimation genetic variance, in particular the additive genetic variance. The contribution of nonadditive genetic variation to phenotypic variation for complex traits in human populations continues to be debated (8, 35, 97) and is still unresolved, mostly because current experimental designs lead to biased and imprecise estimates. When we refer to heritability (b^2), we refer to the proportion of the phenotypic variance that is attributable to additive genetic variance.

To estimate additive genetic variance, we specify an additive statistical model (see below) that superimposes a defined variance structure onto the underlying (unknown) variance composition. In all methods described, we test how well the statistical model summarizes the observed data. It is crucial to recognize that even when the data fit the statistical model well, the conclusions we draw may not be true because the statistical model is a simplified version of the true underlying variance structure. For each method, we discuss whether nonadditive variation can be estimated, what the challenges are, and how potential bias from nonadditive and shared family variance may affect the estimate of the additive genetic variance.

THE MIXED LINEAR MODEL

A cornerstone of the theory and application of quantitative genetics is the linear mixed model (16) of the form

$$y = f + a + e. \quad (7)$$

In this model, y represents the measured quantitative trait or phenotype, f represents known fixed nongenetic variables (such as overall mean, sex, and age), and a and e represent the random additive genetic and residual effects, respectively. Residual effects refer to effects that are not accounted for by the fixed and random effects specified in the model, e.g., measurement error. The model is called mixed because it jointly accounts for fixed (f) and random (a and e) effects and called linear because the



various terms are additive in their effect on the trait.

In matrix notation, the linear mixed model represented in Equation 7 can be generalized as

$$\mathbf{y} = \mathbf{X}\mathbf{f} + \mathbf{Z}\mathbf{a} + \mathbf{e}, \quad (8)$$

where \mathbf{y} is a vector containing the phenotypic values, \mathbf{f} is a vector of fixed effects with incidence matrix \mathbf{X} , \mathbf{a} is a vector of random additive genetic effects with incidence matrix \mathbf{Z} , and \mathbf{e} is a vector of residuals. The elements in the incidence matrices are either zero or one, depending on whether the relevant effect is present in the individual.

Crucially, elements in the vector \mathbf{a} are correlated because individuals share genes by descent from a common ancestor. We define the covariance matrix for the vector \mathbf{a} of genetic effects as \mathbf{G} and the covariance matrix for the vector \mathbf{e} of residuals as \mathbf{R} . The covariance matrix for the vector of phenotypic values \mathbf{y} is then

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}, \quad (9)$$

where the term $\mathbf{Z}\mathbf{G}\mathbf{Z}^T$ represents the variance-covariance matrix attributed to the random genetic effects, and \mathbf{R} represents the variance-covariance attributed to the residual effects. If we assume that the residual effects are independent and have constant variance, \mathbf{R} is a diagonal matrix ($\mathbf{R} = \sigma_e^2\mathbf{I}$). However, this assumption does not hold if there are shared environmental effects between subsets of individuals, such as families. In that situation the general form \mathbf{R} , or an equivalent model that includes the shared environmental effect, must be used.

The statistical and computational analysis of more advanced versions of Equation 8 for large data sets was facilitated by C.R. Henderson (33, 34), who developed efficient algorithms to simultaneously obtain estimators of the fixed effects [best linear unbiased estimator (BLUE)] and predictors [best linear unbiased predictor (BLUP)] of the random effects. In animal and plant breeding, BLUP is widely used to predict the breeding value of individuals in selection programs (25), whereas in human genetics

BLUP could be used to make predictions on disease susceptibility (56).

The main objective of the applications reviewed here is to estimate and partition genetic and environmental variance. When variances are known, the MLM allows the simultaneous estimation of the fixed effects (BLUE) and prediction of the random effects (BLUP). In practice, the variance components are usually estimated using maximum likelihood (ML) or restricted maximum likelihood (REML) (58), and the estimates of the variance components are subsequently used to obtain the BLUE and BLUP estimators of the fixed and random effects. Estimates of variance components from maximum likelihood are biased, in particular when there are many fixed effects in the model relative to the number of observations. REML does not suffer from this bias. In applications to human data sets, the number of fixed effects is usually small relative to sample size and ML and REML estimates are very similar.

The genetic effect (a) for an individual person is the sum of all effects at causal loci in the genome and is assumed to be drawn from a specified distribution, usually a normal distribution. We define a as the additive genetic value of the individual and hence its variance as the additive genetic variance. Note that nonadditive genetic variation is ignored in this model.

The differences between the three methods we discuss are reflected in the genetic relationship matrix \mathbf{G} used in the MLM to estimate genetic variance (Equation 9). In pedigree designs without genetic-marker data, elements of \mathbf{G} are the coefficients of expected genetic relatedness between relatives, derived from the probabilities of identity-by-descent (IBD) on the basis of the recorded pedigree, e.g., in the classical twin design, 1 for MZ twin pairs, 1/2 for DZ twin pairs, and 0 for unrelated individuals. In the within-family design, where the pedigree is known and genetic-marker data are available, elements of \mathbf{G} are the realized or actual coefficients of relatedness, and these coefficients vary around 1/2 for full siblings and are zero for individuals from different families. In the population design in which the pedigree

Breeding value: the sum of the additive effects of an individual's genes. In human genetics we use the term additive genetic value instead

Best linear unbiased predictor (BLUP): A method for prediction of the random effects in a linear mixed model

Identity-by-descent (IBD): situation in which two alleles are inherited from a common ancestor



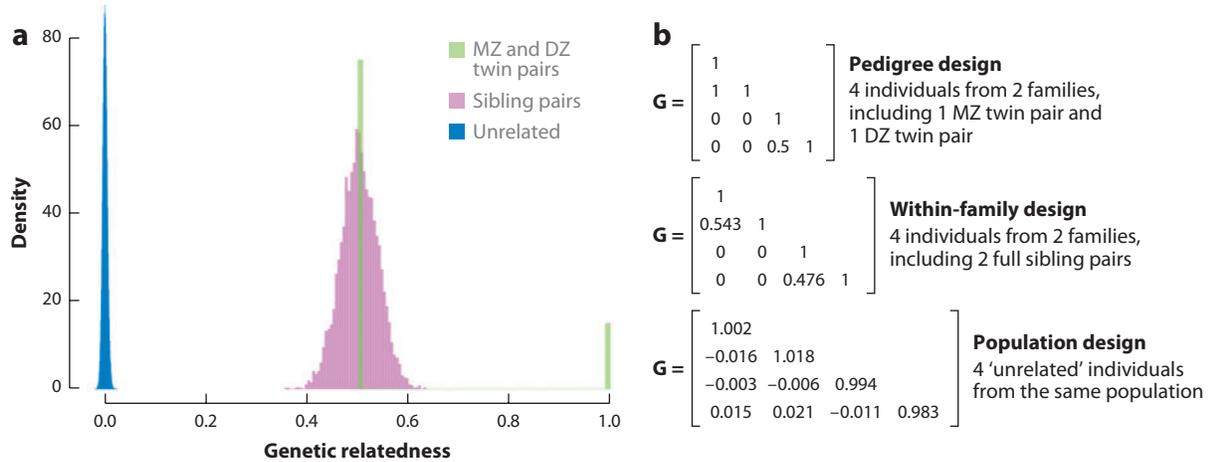


Figure 1

(a) Distributions of the off-diagonal elements for the genetic relationship matrix \mathbf{G} in the pedigree design using monozygotic (MZ) and dizygotic (DZ) twin pairs (green), the within-family design using full-sibling pairs (pink), and the population design using unrelated individuals (blue). The distribution of off-diagonal elements for the genetic relationship matrix \mathbf{G} for the pedigree design represents the expected proportion of genome-wide identity-by-descent (IBD) for MZ (=1) and DZ (=0.5) twin pairs. The distribution of actual proportion of the genome that is shared IBD within full-sibling pairs in the within-family design varies around 0.5, whereas the distribution of the actual proportion of the genome that is shared identity-by-state (IBS) for individuals not knowingly related to each other varies around zero. IBD and IBS estimates for the within-family design and the population design, respectively, are derived from data from the Framingham Heart Study (FHS). Mean IBD was estimated at 0.5028 with a standard deviation of 0.0368 for full siblings. Mean IBS was estimated at -0.0002 with a standard deviation of 0.0046 for pairs of unrelated individuals. Only full-sibling pairs were selected for the IBD estimation in the within-family design, whereas for the IBS estimation in the population design only one sibling per family was selected and one member of each of the remaining pairs of individuals that had an estimated genetic relationship of more than 0.025 was removed. Estimates of IBS are relative to an arbitrary base population with an average relationship between all pairs of individuals of zero. In this analysis, the sample under study is used as the base population; consequently, the average relationship between all pairs of individuals is zero, and the average relationship of an individual with him or herself is one. (b) Examples of the genetic relationship matrix \mathbf{G} for four individuals in the pedigree design (upper matrix), the within-family design (middle matrix), and the population design (lower matrix). Diagonal elements in matrix \mathbf{G} represent an individual's estimated genetic relatedness with him or herself. Off-diagonal elements in matrix \mathbf{G} represent genetic similarity between individuals. Note that matrix \mathbf{G} is symmetrical and that for clarity only elements in the lower triangular are provided. In the pedigree design, off-diagonal elements represent expected genome-wide IBD sharing for four individuals (one MZ twin pair and one DZ twin pair) from two independent families in which parents are assumed to be unrelated. In the within-family design, the off-diagonal elements represent actual variation around the expected genome-wide IBD sharing [$E(\hat{\pi}) = 0.5$] for two independent full-sibling pairs. In the population design, the off-diagonal elements represent actual genome-wide IBS for four (i.e., six pairs) individuals that are not knowingly related. In the population design, an individual's genetic relatedness with him or herself (diagonal elements) is an estimate of $1 + F$, with F being the inbreeding coefficient relative to the base population.

is unknown but dense genetic-marker data are available, \mathbf{G} contains estimates of coefficients of additive genetic covariance between pairs of individuals that are captured by the markers used to construct \mathbf{G} . These coefficients are scaled to vary around zero for pairs of individuals that are not knowingly related. See **Figure 1b** for graphical representation of matrix \mathbf{G} for the three designs.

In all the designs we review, the sampling variance of the estimate of heritability is a function of sample size and the variation among the elements of \mathbf{G} ; more variation implies smaller sampling variance (see **Figure 1a** for a graphical representation of the distribution of elements in matrix \mathbf{G}). In extended or complex pedigrees, the coefficients in matrix \mathbf{G} are $1/2^k$ for individuals and descendants who are



Table 1 Precision of estimates of heritability from whole-genome methods

Design	Sampling variance of estimate	References
Pedigree design including an equal number of MZ and DZ twin pairs	$\text{var}(\hat{b}^2) \approx 4 \left(\frac{1}{m} + \frac{1}{m} \right) = 4 \left(\frac{2}{m} \right) = \frac{8}{m} = \frac{32}{N}$	NA
Within-family design including full sibling pairs	$\text{var}(\hat{b}^2) \approx \frac{2}{m * \text{var}(\hat{\pi})}$ $\text{var}(\hat{\pi}) \approx 0.0015$ $\text{var}(\hat{b}^2) \approx \frac{4}{N * 0.0015} \approx \frac{2667}{N}$	(65, 83, 85)
Population design including only distantly related individuals (e.g., pair-wise genetic relationship <0.025)	$\text{var}(\hat{b}^2) \approx \frac{2}{N^2 * \text{var}(\hat{\pi})}$ $\text{var}(\hat{\pi}) \approx \frac{\log(4 * N_e * L)}{(2 * N_e * L)} \approx 2 \times 10^{-5}$ $\text{var}(\hat{b}^2) = 1 \times 10^5 / N^2$	(68, 82) (25)

Abbreviations: DZ, dizygotic; b^2 , heritability estimate; L , total map length over all autosomes ($L = 35$) (41); m , number of MZ, DZ, or full sibling pairs; MZ, monozygotic; N , number individuals; $\hat{\pi}$, estimated proportion of genome-wide identity-by-descent; N_e , effective population size and is assumed to be 10,000; var, variance.

k generations apart. Hence, sampling variation in a (human) pedigree design using close relatives is small compared with designs using more distant relatives. Bias, however, is more likely to come with analyses of close relatives and is generally of more concern than precision. To obtain tractable and comparable approximations, we have assumed that the population value of heritability is small (i.e., < .40). When the true population value is large (e.g., as it is for height: $b^2 = .80$) (70, 85), the actual sampling variance is smaller than our approximation. **Table 1** summarizes the precision of the estimates of genetic parameters for all methods.

ESTIMATING HERITABILITY FROM EXPECTED GENETIC RELATEDNESS OF RELATIVES IN A PEDIGREE DESIGN

Design

In pedigree studies, genetic parameters are estimated from phenotypic similarity between known relatives (21, 49). Twin studies are a special case of pedigree studies and have been used to estimate heritability for a wide variety of traits, including disease susceptibility, anthropometric traits, and behavioral phenotypes (6).

Heritability estimates vary widely (0 to 0.8), but for many traits heritability is estimated as moderate to high (in the range of 0.4–0.8).

In the pedigree design, the phenotypic resemblance [often denoted as the intraclass correlation (t)] of MZ and DZ twin pairs is utilized to estimate the contribution of genetic and environmental variation to the phenotypic variation of a trait. MZ twins share 100% of their genomic variation IBD, whereas DZ twins share on average 50% of their genome IBD. Hence, the matrix **G** has pair-wise coefficients of 1 (MZ twins), 1/2 (DZ twins), and 0 (individuals from different families). Assuming that the common environmental variance is equal for MZ and DZ twins, narrow-sense heritability can be estimated as twice the difference between phenotypic correlations for MZ and DZ twin pairs (21). In practice, maximum likelihood methods are used to estimate (co)variance components (e.g., see 54), and the MLM (Equation 8) is augmented with additional random effects, e.g., the effect of a shared environment or nonadditive genetic effects.

Precision

An approximate asymptotic expression for the sampling variance of the estimate of

Intraclass correlation: the correlation between measures within a group (e.g., MZ or DZ twins)

Narrow-sense heritability: proportion of the phenotypic variance that is explained by additive genetic variance



Annu. Rev. Genet. 2013.47. Downloaded from www.annualreviews.org by University of Queensland on 09/02/13. For personal use only.

Broad-sense heritability:

proportion of the phenotypic variance that is explained by both additive and nonadditive genetic variance

narrow-sense heritability from the classical twin design, assuming an equal number (m) of MZ and DZ twin pairs, is $4(1/m + 1/m) = 32/N$ (Table 1) (80), with N being the total number of individuals with a measured phenotype ($N = 4m$). For example, for 200 MZ and 200 DZ pairs (800 samples in total), the approximate standard error of the estimate of heritability is $\sqrt{(32/800)} = 0.2$.

Limitation and Potential Bias

A limitation of the classical twin design (with MZ and DZ twin pairs) is that it allows estimation of only three variance components (including residual variation) because there are only two estimates of correlation (one for MZ pairs and one for DZ pairs) from which variance components are estimated. Consequently, if the true population variance contains more than three sources of variance, they cannot all be estimated simultaneously in the model. For instance, if MZ twins share a more similar environment than DZ twins, this effect cannot be separated from the increased genetic similarity between MZ twins compared with DZ twins. Similarly, nonadditive genetic effects decrease the correlation for DZ twins relative to the correlation of MZ twins and therefore may lead to overestimation of the narrow-sense heritability, whereas shared environmental effects increase the correlation for DZ twins relative to the correlation of MZ twins and may lead to underestimation of the broad-sense heritability. Limitations of this design lead to strong assumptions about causes of family resemblance, e.g., absence of nonadditive genetic variance (21, 49, 97), and consequently broad-sense and narrow-sense heritability cannot reliably be distinguished. Extending the classical twin design with other relatives, such as parents, spouses, and (adopted) children of the twins allows a wider range of models to be fitted and allows testing of some, but not all, assumptions (e.g., 40, 51, 79). However, collection of large cohorts is difficult.

Example for Human Height

In a comparative study on human height measured in Caucasian twin cohorts from eight different countries, Silventoinen et al. (70) estimated heritability from MZ and DZ twin pair resemblance. Data were available for 30,111 complete pairs. MZ twin correlations ranged from 0.87 to 0.94 in both men and women, whereas same-sex DZ twin correlations ranged from 0.42 to 0.57 in men and from 0.49 to 0.56 in women. Opposite-sex DZ twin correlations ranged from 0.30 to 0.50 but were not included in the modeling. Maximum likelihood was used to estimate the contribution of genetic and environmental effects. The contribution of shared environmental factors was generally low and nonsignificant in most cohorts. Heritability estimates ranged from 0.70 to 0.87 in men and from 0.68 to 0.93 in women. Although substantial variation was observed in mean body height across different cohorts, the relative contribution of genetic factors was very similar between populations. Both the observed resemblance between close relatives, including parents and offspring, and the inference on heritability for human height have been consistent for more than a century (84). Empirical observations on the resemblance between twins and other close relatives suggest that most phenotypic variation in human height in the population is genetic and that most genetic variation is additive.

ESTIMATION OF HERITABILITY FROM ACTUAL GENETIC RELATEDNESS IN A WITHIN-FAMILY DESIGN**Design**

In pedigree studies (e.g., twin and other family studies), additive genetic variance is estimated from expected genome-wide IBD sharing between relatives. These studies are based on strong assumptions about the covariance between individuals within and between families. For the estimation of additive genetic

100 Vinkhuyzen et al.



variation, these assumptions can be bypassed by utilizing only within-family information if very large data sets are available.

Through Mendelian segregation, full siblings of non-inbred parents share zero, one, or two copies of the alleles at each autosomal locus, with probabilities of 1/4, 1/2, and 1/4, respectively. The total expectation of IBD in a population then becomes 1/2 with variance of 1/8 for a single locus. The variance becomes smaller when the number of loci increases (30, 36, 64), with the expected proportion of IBD sharing being equal to the actual proportion of IBD sharing if genetic variance were due to an infinite number of independent loci. However, the number of loci is limited, and genetic linkage causes dependent segregation of loci in the pedigree, maintaining variation around the expected genetic similarity for all pairs of relatives, apart from MZ twin pairs (who always share both alleles IBD) and parents and offspring (who always share one allele IBD). Using genetic-marker data, we can precisely estimate the amount of the genome shared by a pair of relatives and can estimate narrow-sense heritability by simply regressing phenotypic similarity on their genome-wide genetic similarity.

Using information from exactly two full siblings per family, the matrix \mathbf{G} in this design is block-diagonal and contains off-diagonals that are estimates of the realized or actual proportion of the genome that is shared IBD (π_G) for a pair of siblings. Elements for individuals from different families are zero (**Figure 1a**). Estimates of locus or genome-wide IBD are obtained from genetic-marker data, and efficient algorithms exist to calculate probabilities of IBD, in particular for small pedigrees (e.g., see 1, 17, 43).

For analysis of the within-family design, Equation 8 is augmented by a random effect that models the covariance between full siblings that is the same for all pairs; irrespective of the proportion of their genome, they share IBD. This random effect is included because the objective of the analysis is to estimate additive genetic variation from the deviation of the

expected value of IBD, which is 1/2 for all pairs. The covariance between sibling pairs i and j is

$$\text{cov}(y_i, y_j) = \pi_{Gij}\sigma_A^2 + \sigma_C^2, \quad (10)$$

where σ_A^2 is the additive genetic variance and σ_C^2 represents the sibling covariance not explained by additive genetic effects, such as effects of dominance, epistasis, and the shared family environment. Because the heritability in this design (85) is estimated solely from segregation within families without any assumptions regarding underlying factors that cause between-family variance, the estimate of additive genetic variance is free of confounding by environmental differences between families. As is the case for the pedigree design, the estimate of additive genetic variance from the within-family design is biased upwards when there are nonadditive genetic effects and when their variance is not estimated simultaneously with the variance of additive effects.

Precision

The sampling variance of the estimate of narrow-sense heritability of this design is approximately $2/(m \cdot \text{var}(\pi_G))$, with m being the number of full-sibling pairs and $\text{var}(\pi_G)$ being the variance in pair-wise-realized genetic relationships between the siblings. Theoretical studies provide an approximation of variation in realized relationships of full-sibling pairs: $\text{var}(\pi_G) \approx 1/(16L) - 1/(3L^2)$, where L is the total length of the genetic map (in Morgans) (85). For humans, the total map length for the 22 autosomes is $L \approx 35$ (41), and so $\text{var}(\pi_G) \approx 0.039^2$, which is close to what has been reported empirically (65, 83, 85). Hence, around the expected proportion of shared IBD (i.e., 1/2), there is a standard deviation of roughly 4%. Because the number of recombination events per chromosome is small (27), genome-wide IBD sharing between full siblings can be estimated with only a few markers per chromosome. The approximate sampling variance of the estimate of heritability of the full-sibling design is $2667/N$, given a number of assumptions, with N being



Genome-wide association study (GWAS): a study design in which hundreds of thousands of genetic variants in the genome (usually SNPs) are tested for association with the trait

Linkage disequilibrium (LD): a measure of association between alleles at different loci

the total number of individuals with a phenotype (**Table 1**).

Limitation and Potential Bias

The limitation of the within-family design is the large sample size required to estimate parameters with sufficient precision. For the same number of people with a phenotype, this design is approximately 80 times less efficient than the twin design with equal number of MZ pairs as DZ pairs. For 10,000 full-sibling pairs (20,000 phenotypes), the approximate standard error is 0.37. Potential bias may come from nonadditive genetic effects that are not modeled because limited sample size does not allow a reliable distinction between additive and non-additive effects. Both additive and nonadditive genetic variance are estimated from variation around the expected coefficient of relatedness; for full siblings, the expected coefficient of relatedness is 1/2 for additive genetic variance and 1/4 for dominance genetic variance and additive \times additive genetic variance (21). In theory, this design allows estimation of genetic dominance deviation. However, because the coefficients for additive variance and dominance variance are highly correlated [theoretical value is 0.89 (85)], a strong sampling correlation between the estimates is expected, which implies that even larger sample sizes are required to reliably distinguish nonadditive from additive genetic variance.

Example for Human Height

In a study on human height measured in 3,375 quasi-independent sibling pairs, Visscher et al. (85) estimated heritability by correlating phenotypic similarity and genome-wide IBD sharing between siblings. Actual genome-wide IBD for full siblings ranged from 0.374 to 0.617 with a mean of 0.498 (with a standard deviation of 0.036). Two models were fitted: a full model that included a genome-wide additive effect, a shared environmental effect, and a residual effect, and a reduced model that excluded the genome-wide additive effect. Maximum like-

lihood was used to estimate the contribution of genetic and environmental effects using the MLM as previously described. Heritability was estimated at 0.80 (95% confidence interval; 0.46–0.85), with the remaining variance completely attributable to the residual effect, which is very similar to estimates and inference from twin and family studies (70). The within-family design, however, facilitated a complete separation of genetic and environmental factors and can therefore be seen as an independent validation study for estimating the heritability of human height from pedigree studies. A subsequent study with more than 10,000 full-sibling pairs reported a similar value and also partitioned additive genetic variation into contributions from individual chromosomes (83). Results showed that the data were consistent with a model in which variance explained by a chromosome is proportional to the length of that chromosome.

ESTIMATING HERITABILITY FROM POPULATION-BASED ESTIMATES OF GENETIC RELATEDNESS

Design

Advances in genotyping technologies have led to arrays of SNPs that can genotype hundreds of thousands to millions of markers in a single assay (3). These SNP chips form the basis of genome-wide association studies (GWASs), which have revolutionized human genetics in the past six years (e.g., see 87). **Table 2** shows GWAS results for a selection of quantitative traits.

In a GWAS, association between each SNP and the trait is tested, and the paradigm is based on the existence of linkage disequilibrium (LD) between ungenotyped causal variants and SNPs in the analysis. LD refers to a nonrandom assortment of alleles at two loci and occurs in a finite, random mating population because chromosomal segments are descended from a common ancestor without any recombination. Consequently, chromosomes that carry the same allele at a locus that affects a complex

102 Vinkhuyzen et al.



Table 2 Proportion of variance explained by genetic factors for a number of selected quantitative traits

Trait	b^2 pedigree design ^a	b^2 GWAS hits ^b	b^2 population design ^c
Height	0.80 (70)	0.10 (94)	0.45 (91, 95)
Body mass index	0.45–0.80 (67)	0.02 (73)	0.17 (95)
von Willebrand factor	0.66–0.75 (12, 57)	0.13 (71)	0.25 (95)
Bone mineral density	0.61 (2)	0.06 (18)	0.16 (93)
General intelligence			
- Children (~12 years)	0.40–0.60 (4, 32)	0 (5)	0.22–0.64 (5)
- Adults	0.80 (32, 61)	0 (9)	0.40–0.50 (11)
Red blood cell phenotypes			0.16 (93)
- Hemoglobin concentration	0.84 (19)	0.02 (76)	
- Sodium	0.50 (88)	0.02 (93)	
Personality			
- Neuroticism	0.13–0.58 (38)	0 (13)	0.06 (78)
- Extraversion	0.34–0.57 (38)	0 (13)	0.12 (78)

^aHeritability in the pedigree design is estimated by comparing expected and observed monozygotic and dizygotic twin pair resemblance.

^bHeritability from genome-wide association study hits represents the total variation explained by all of the single nucleotide polymorphisms that individually reached genome-wide significance in genome-wide association studies.

^cHeritability in the population design is estimated from the single-nucleotide-polymorphism-derived genetic similarity between pairs of individuals that are not knowingly related.

trait are also likely to carry the same allele at a nearby SNP, generating an association between the SNP alleles and the trait (55).

Because of the large number of tests conducted in a genome-wide survey, very stringent type-I error rates are used [$\sim 5 \times 10^{-8}$ (14, 26)] to avoid false positives and to ensure that reported associations are robust and are likely to replicate in other samples from the same population. This stringent threshold minimizes false positives but leads to many false negatives because a causal variant with a small effect or weak LD with SNPs on the chip does not generate an association between any one SNP and the trait that is large enough to be declared significant. In practice, the effect of these false negatives has been found to be dramatic. For most traits, the SNPs that are declared significant explain in total 10% or less of the genetic variance. This has been referred to as the “missing heritability” paradox (52).

Instead of testing the effect of each SNP independently on the trait, it is possible to estimate the variance explained by fitting all the SNPs simultaneously. This is equivalent to estimating the relationship between individuals

from the SNPs and using this relationship matrix to estimate the genetic variance (25, 74, 77). A method described by Yang et al. (91, 92) utilizes LD between genotyped SNPs and unknown causal variants to capture additive genetic variation underlying phenotypic variation in a random sample of unrelated individuals in the population. In this design, matrix \mathbf{G} represents genetic similarity between individuals j and k from m genotyped SNPs:

$$G_{jk} = \frac{1}{m} \sum_{i=1}^m \frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1 - p_i)}, \quad (11)$$

where p is the frequency of the reference allele and x_i is the genotype indicator of the i^{th} SNP ($x_i = 0, 1, \text{ or } 2$). Estimates of genetic similarity are the genetic relationships expressed relative to a base population; in this method, the study sample is the base (whereas in pedigree studies, the base is the set of founders with no recorded or inferred relationships to older individuals). In the equation above, the average similarity is zero if the allele frequencies (p) are estimated from the sample because the expected value of x is $2p$. This is also the matrix that is used for

Type-1 error: rejection of the null hypothesis when the null hypothesis is true



principal component analysis to infer population structure from SNP data (62).

The basic idea behind this method is to estimate additive genetic variance by including all the SNPs in the model without focusing on individual SNPs. In other words, it is an estimation rather than a hypothesis-testing paradigm. Variance explained by causal variants that are in LD with genotyped SNPs, but whose effect sizes are too small to reach genome-wide significance in a conventional GWAS, are included in the heritability estimate derived through this method.

Yang et al. (91) showed that estimates of additive genetic variation using this method directly address the perceived problem of missing heritability (52, 53). Estimates of additive genetic variation quantify how much variation is captured by all SNPs, and therefore quantify how much variation would be explained by a GWAS when the sample size is so large that all variants that are associated would be statistically significant (86, 91, 92).

Precision

As mentioned above, sampling variance increases with decreasing variation among the coefficients of \mathbf{G} , assuming all else is equal. With expected coefficients being $1/2^k$ (for k^{th} degree relatives), the population design with only distant relatives yields only little variation. However, whereas \mathbf{G} in the pedigree design is block diagonal with coefficients of expected or realized IBD for within-family pairs and coefficients of zero for all other elements representing pairs of unrelated individuals, \mathbf{G} in the population design is filled with estimates for all pairs of individuals. The number of pairs in the population design is $(N^2 - N)/2$, which is a multiple of the number of pairs in the pedigree design ($= 1/2 N$ for pairs of twins). The precision in the population design comes from the very precise estimate of genetic similarity and from the large number of pair-wise comparisons in the sample.

Theory borrowed from linkage analysis of quantitative traits (68, 82) predicts that the sam-

pling variance of the estimate of heritability from the mixed model analysis is approximately $100,000/N^2$ (Table 1), hence a standard error of $315/N$. We validated this by simulations using GWAS data of the Atherosclerosis Risk in Communities Study (ARIC) cohort (63, 95). The simulation results are provided in the supplemental material (follow the **Supplemental Material link** from the Annual Reviews home page at <http://www.annualreviews.org>).

Limitation and Potential Bias

A limitation of this design is that genetic variance contributed by causal variants that are not in sufficient LD with the genotyped SNPs is not included in the heritability estimate. If LD between the genotyped SNPs and the causal variants is incomplete, the genetic similarity between individuals j and k at the causal variants is different from the genetic similarity between those individuals estimated from the genotyped SNPs (89). Consequently, genetic variance from untagged causal variants is not accounted for by the genetic similarity calculated from genotyped SNPs. This is where the matrix \mathbf{G} differs from the previous applications. In the pedigree and within-family designs, the coefficients of relatedness are based on IBD and are blind to allele frequencies of causal variants (if DNA segments in a pair of individuals are IBD, then any variant in that segment, common or rare, is shared). In contrast, the population-based estimate of SNP sharing relies on LD and is sensitive to allele frequencies. In the extreme case that all causal variants in the genome are at low allele frequency in the population, and therefore not in LD with common variants (89), the pedigree and within-family design estimates total heritability, whereas the SNP-based estimates is zero (and a GWAS would not work either). Recently, Speed et al. (72) proposed a method in which the contribution of the SNPs to the estimate of genetic similarity between a pair of individuals is weighted according to the LD with their neighboring SNPs, aiming to reduce potential bias and to increase the precision of the heritability estimate.



Bias may also come from a shared environment that is not modeled in this design. If individuals who share SNP genotypes more often than the average also tend to share a common environment, then the heritability explained by the SNPs will be overestimated. This would be expected if closely related people (e.g., cousins, siblings) were included in the sample. However, if closely related people are excluded from the analysis, this source of bias should be small because among distantly related people, genomic similarity is poorly correlated with pedigree relationship, and it is only the pedigree relationship that might be correlated with environmental similarity.

Another possible source of bias may arise if the population consists of subpopulations that differ both genetically and environmentally. This bias is usually avoided by testing for population structure and eliminating it from the data or correcting for it in the analysis (e.g., by including the first few principal components of the relationship matrix as fixed effects in the MLM). Biases attributable to population structure and genotyping artifacts, such as plate and batch effects, are more likely to be a problem in case-control analyses than quantitative trait analyses because confounding with the binary phenotype is not uncommon (44, 45).

Example for Human Height

Yang et al. (91) estimated the heritability of human height from 294,831 SNPs genotyped on 3,925 unrelated individuals. The data were fitted in an MLM, and REML was used to estimate the variance explained by the SNPs. All the SNPs were considered simultaneously in the model and the proportion of phenotypic variance explained by the SNPs was 0.45 (SE 0.08), with remaining variance due to nonfamilial environmental factors and possible measurement error. This estimate of genetic variance forms the lower boundary of total narrow-sense heritability because only genetic variation due to causal variants that are in sufficient LD with the genotyped SNPs is included in the estimate. Using simulated data,

the authors show that incomplete LD between the causal variants and the genotyped SNPs can explain all of the remaining heritability. This study showed that the so-called “missing heritability” (52) for height is not missing but hidden. The finding that, together, common variants explain a substantial proportion of the heritability proves that GWASs to date have been underpowered to detect individual SNPs with small effects. Results from this study indicate that much larger sample sizes are required to detect those individual common SNP effects in a GWAS. The observed strong relationship between the experimental sample size of a GWAS and the number of significant loci detected (81) is consistent with that conclusion. Causal variants in low LD with the genotyped SNPs may be identified through whole genome sequencing studies.

A SUMMARY AND COMPARISON OF THE METHODS

We have reviewed three different methods to estimate heritability from genetic similarity between pairs of relatives. The methods differ in the parameters estimated. The use of the twin (pedigree design) and full-sibling (within-family design) data leads to an estimate of total heritability, whereas the use of SNPs to construct a relationship matrix (population design) estimates the additive genetic variance (and therefore heritability) explained by the SNPs. The proportion of genetic variance explained by the SNPs depends on the structure of the data. In a population of unrelated individuals, this proportion depends on the LD between SNPs and causal variants. Although individuals are not known to be related, they could share distant ancestors and therefore some chromosome segments that are IBD. Thus, the estimate of genetic variance from all SNPs can be considered as driven by LD or by distant realized relationships: The two descriptions are equivalent. In fact, the variance of the SNP-based relationship is equal to the LD averaged over all pairs of SNPs. The within-family design, which uses variation in realized relationships between

pairs of SNPs, can be described as using LD generated by the inheritance of large chromosome segments from parents to offspring within a family. Consequently, this design estimates the full heritability because in this design the SNPs track all causal variants.

Heritability estimates derived through these designs differ in precision and possible bias. Generally, close relatives give more precision but potentially more bias, whereas distant relatives give less precision and less bias. Bias in analyses of close relatives may come from environmental variation that is confounded with additive genetic variation within families, or in the case of siblings, confounded with nonadditive genetic effects. Precision in parameter estimates depends on the total number of individuals with a phenotype and the variation in relationship (close or distant).

The heritability estimated in the pedigree design has generally small sampling variance, as much variation exists among the coefficients of relatedness. In this design, however, potential bias is of greater concern and generally has a larger impact on the estimate than precision. The heritability estimated from variation around the expected genetic similarity of full siblings in a within-family design is free of assumptions about variation between families, but relative to the classical twin design, the sampling variance is large. Inflation of the estimate of narrow-sense heritability by nonadditive genetic variation is a concern in both the pedigree and the within-family design. In the pedigree design, this is caused by confounding with environmental factors. In the within-family design, this is caused by a strong correlation of sampling variance between additive and nonadditive genetic variation. Consequently, the power to estimate nonadditive genetic variance is generally insufficient. Heritability estimated in the population design is unbiased: Distant relatives are unlikely to share variation due to environmental factors or variation due to nonadditive effects. Sampling variance is relatively small due to a large number of pairwise comparisons that can be achieved with the large sample sizes that are available to date.

The population design is as efficient as a twin design when tens of thousands of individuals with phenotypic and SNP data are available. Estimating heritability from genetic similarity of distant relatives requires smaller sample sizes compared with the within-family design to obtain similar precision. However, a much larger number of genetic markers is required to accurately estimate the genetic similarity of distant relatives in the population design. SNP chips available to date are adequately designed to ensure sufficient LD between genotyped variants and ungenotyped causal variants. Hence, heritability can be estimated from distant relatives without much error, especially when heritability is high.

GENETIC ARCHITECTURE

Numerous pedigree studies have revealed moderate-to-large heritability estimates for a wide variety of complex traits in human populations. After the completion of the Human Genome Project, considerable success of GWASs was anticipated. After six years of GWAS discovery, however, much of the genetic variance estimated from pedigree studies has not been accounted for by the genetic variants discovered from GWASs. For complex traits, typically less than 10% of the genetic variation is explained by SNPs, although there are exceptions: For age-related macular degeneration, an eye disease, approximately 50% of genetic variation has been accounted for by only five loci (31), and for Crohn's Disease and ulcerative colitis, two inflammatory bowel diseases, very large experimental sample sizes (~15,000 cases) have led to the discovery of hundreds of loci by GWASs, which in combination with known less common variants explain approximately 20% of genetic variation (39).

Several explanations have been raised to answer the case of the missing heritability (52). Possible explanations are that (a) pedigree studies have overestimated the heritability by, for example, bias due to nonadditive and/or environmental effects, that (b) causal variants individually explain such a tiny amount of

variation that their effects do not reach statistical significance in GWASs to date, and/or that (c) causal variants are not in sufficient LD with the genotyped SNPs and therefore their effects are not fully captured by the genotyped SNPs in GWASs. Quantifying the difference between (b) and (c) is informative with respect to the allelic spectrum of causal variants, i.e., the frequency of risk alleles in the population. Causal variants that are in low frequency in the population are not in high LD with genotyped SNPs. Consequently, variation caused by variants that are not in sufficient LD with SNPs cannot be captured and therefore remains undetected in GWAS.

Whole-genome methods utilizing an expected or realized genetic relationship between individuals have increased our understanding of the genetic variation that underlies complex traits. A short recapitulation for human height, pedigree studies using data from MZ and DZ twin pairs have repeatedly reported heritability estimates of approximately 80% (e.g., 50, 70), an estimate confirmed by Visscher et al. (83, 85) utilizing empirical genome-wide IBD sharing of full-sibling pairs (the within-family design). Using realized genetic similarity between unrelated individuals, Yang et al. (91, 95) have demonstrated that 45% to 55% of the phenotypic variance can be explained by common SNPs when taking the individual SNP effects together. The largest GWAS on height to date (42) has identified 180 genetic loci that together explain ~10% of the phenotypic variation. Together, these results suggest that for human height the genetic variance is additive and involves many loci of small effect. The difference between ~50% and 10% is due to SNP associations with height that are too small to reach the stringent significance level used in GWASs. Individual loci do not explain much of the genetic variation because otherwise these effects would have been identified by GWASs to date. The variance unaccounted for (80% – 50% = 30%) is likely to be mainly due to the segregation of causal variants at low frequency. Bias due to nonadditive genetic effects or environmental variation seems unlikely for

human height given the similar heritability estimates derived from different study designs and additive genetic models that fit the data well in all three study designs. Nevertheless, the power to test for absence of nonadditive variance in any of the reviewed designs remains low so even if the data are consistent with a parsimonious additive model, it doesn't prove that there is no variation due to dominance of epistasis.

We summarize the proportion of variance explained from pedigree analyses, genome-wide significant SNPs, and population-based analyses using estimated pair-wise genomic relationships for a selection of quantitative traits in **Table 2**. Traits other than height, such as body-mass index and cognitive ability, follow the same trend, i.e., the estimates of heritability from pedigree designs is large, genome-wide significant loci explain none or a small proportion of phenotypic variation, and variance estimated from genomic relationships captures 1/3 to 2/3 of pedigree heritability. For height, heritability estimates derived from pedigree studies have been confirmed by other study designs, and biased estimates are unlikely. For other traits, however, inflated (e.g., due to nonadditive genetic variation) estimates from pedigree studies are a possible cause of part of the missing heritability. The contribution of nonadditive genetic variation to phenotypic variation for complex traits in human populations continues to be debated (8, 35, 97) and is unresolved, mostly because current experimental designs lead to biased and imprecise estimates.

DISCUSSION

One of the aims of quantitative genetics has been to quantify the amount of variation in complex traits that is due to genetic variation and the amount due to environmental variation. This is difficult in humans because people who share genes also tend to share environments. Traditional designs, such as the pedigree design, at least partly overcome this problem. The availability of genome-wide SNP data has allowed the use of new designs, such

as the within-family design through which we can estimate the genetic variance underlying the trait without the confounding of the environment. Within-family studies have tended to confirm traditional estimates for height but their low power means that the standard errors on estimated heritability have been high. Consequently, this design was infeasible for many traits for which sample sizes were too small. Designs based on a population sample of unrelated people can also overcome the confounding of genes and environment but only estimate the proportion of genetic variance explained by the SNPs, which is typically 1/3 to 2/3 of the traditional estimate.

Population designs to estimate genetic variation in a two-stage procedure, by first estimating relatedness from genetic markers and then estimating heritability by contrasting genetic similarity to phenotypic similarity, are not new. Ritland (66) proposed this for studies in natural populations in which obtaining pedigree information may be impossible. The initial idea of this design was to detect IBD between close relatives (e.g., full siblings and half siblings) from a small number of markers and then correlate estimated relatedness with phenotypic covariance. Conceptually, this approach is more similar to the pedigree design, whereby the pedigree is inferred from IBD sharing of large chromosome segments, than the within-family and population designs we have discussed. However, there is no fixed point at which an inferred pedigree design becomes a population design that relies on LD: Increasing marker density allows the estimation of more distant relationships at the expense of a potential loss of information due to imperfect LD between the markers used to infer relatedness and causal variants for the trait.

The new (population-based) whole-genome methods have shown that numbers of genetic variants with small effect explain a substantial proportion of the heritability for complex traits. These common variants account for the difference between the heritability explained by GWAS hits and the heritability estimated from all of the SNPs in the population design (see

Table 2). From this, we can conclude that large sample sizes lead to detection of more individually significant SNPs. Augmenting SNP genotypes with genome sequence data should help to find the remaining missing heritability, i.e., the difference between the heritability estimates from pedigree studies and the heritability estimated from SNPs. In particular, sequence data should be more powerful where causal variants are rare and hence not in high LD with any SNPs on the SNP chip. Although the individual effect size of such a rare variant can be large, each of these variants is expected to explain a small amount of variance simply because they are rare.

Population-based methods can also be used in a multivariate setting to estimate genetic covariance that can be captured by all the SNPs. In the pedigree and within-family designs, genetic covariance is usually estimated by partitioning phenotypic covariance on individuals who have multiple phenotypic measurements. In the population design, however, genetic covariance can also be estimated from SNP data on unrelated individuals that have only one of the phenotypes measured. That is, genetic covariance can be estimated from two or more independent groups of unrelated individuals, with each group being measured on a different phenotype. Coefficients of the SNP-based GRM are the expected genetic covariance between a pair of individuals, and this covariance can be for the same trait or for different traits. Multivariate analysis using SNP data facilitates the quantification of pleiotropy among complex traits and might be of particular interest for traits that cannot be measured on the same individual, such as two different diseases (e.g., see 46, 47). Another goal of research on the genetics of complex traits is to identify individual causal variants and to elucidate their biological mechanisms. GWASs have identified many genes and even some causal sites within these genes that contribute to variation in particular traits (e.g., see 15, 29, 37). Further research will no doubt identify more individual causal variants. Given that the variance explained by many individual causal variants is so small, it

may never be possible to identify all of them. A slightly different aim is to describe the properties of causal variants as a class rather than to identify them all individually. For example, we would like to know the joint distribution of allelic effects and allele frequencies, and to be able to infer this relationship not all causal variants need to be known. This joint distribution is of interest because it reflects the evolutionary processes underlying complex trait variation, including natural selection, the distribution of effect sizes of new mutations, and past and present population size. Moreover, under-

standing the allelic spectrum of causal variants allows us to provide realistic predictions of the number of genetic variants that are present in the genome and their effect sizes, and hence the sample size that is required to detect the variants. That is, knowledge on the joint distribution of effect size and allele frequency leads to useful information on experimental designs to further dissect complex trait variation. The whole-genome methods reviewed here will bring us closer to this goal and will help in the identification of individual causal variants.

SUMMARY POINTS

1. Expected or realized genetic similarity between relatives (either close or distant) can be used to estimate heritability of complex traits in human populations. Estimates based on close relatives generally yield high precision but may come with bias because of strong assumptions that are violated. Estimates based on distant relatives generally have less bias but lower precision.
2. Whole-genome methods increase our understanding of the genetic variation underlying complex traits in humans. These methods have shown that a substantial proportion of genetic variation is additive and that 1/3 to 2/3 of additive genetic variation is captured by common SNPs.
3. Whole-genome methods have shown that the postulated genetic architectures involving only rare variants are not consistent with the data for many complex traits (see also **Table 2**).

FUTURE ISSUES

1. For many complex traits, most of the heritability is hidden rather than missing. Empirical data analyses and simulation studies suggest that future gene-mapping endeavors should focus on both common and rare variants.
2. Whole-genome methods utilizing a realized genetic relationship as reviewed here could be exploited for the use of whole-genome sequencing data to unravel the combined effects of rare variants underlying complex traits.
3. GWAS data that are widely available to date allow investigation of genetic pleiotropy between different traits (and/or diseases). Whereas in the pedigree design two correlated traits were generally measured in the same individuals, the population design also allows estimation of genetic correlation when the two traits are measured in different individuals, thereby increasing flexibility and possibilities in the study.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract Number N01-HC-25195). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI. Funding for SHARe Affymetrix genotyping was provided by NHLBI Contract N02-HL-64278. SHARe Illumina genotyping was provided under an agreement between Illumina and Boston University. The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by grant number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The authors would like to thank Gibran Hemani for the IBD estimation of the Framingham sibling pairs and Bill Hill for his valuable comments on an earlier version of this review. This research was supported by the Australian Research Council (FT0991360, DP1093502, and DP130102666), the Australian National Health and Medical Research Council (APP1011506, APP1047956, APP1048853, and APP1052684), and the National Institutes of Health (GM099568, GM075091, and MH100141).

LITERATURE CITED

1. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin: rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* 30:97–101
2. Arden NK, Baker J, Hogg C, Baan K, Spector TD. 1996. The heritability of bone mineral density, ultrasound of the calcaneus and hip axis length: a study of postmenopausal twins. *J. Bone Miner. Res.* 11:530–34
3. Barnes MR. 2003. Human genetic variation: databases and concepts. In *Bioinformatics for Geneticists*, ed. MR Barnes, Gray, I.C. Harlow, pp. 459–69. Essex, UK: Wiley & Sons
4. Bartels M, Rietveld MJ, Van Baal GC, Boomsma DI. 2002. Genetic and environmental influences on the development of intelligence. *Behav. Genet.* 32:237–49
5. Benyamin B, Pourcain B, Davis OS, Davies G, Hansell NK, et al. 2013. Childhood intelligence is heritable, highly polygenic and associated with FBNP1L. *Mol. Psychiatry* doi: 10.1038/mp.2012.184
6. Boomsma D, Busjahn A, Peltonen L. 2002. Classical twin studies and beyond. *Nat. Rev. Genet.* 3:872–82
7. Brownlee J. 1911. The inheritance of complex growth forms, such as stature, on Mendel's theory. *Proc. R. Soc. Edinb.* 11:251–56
8. Carlborg O, and Haley CS. 2004. Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* 5(8):618–25
9. Chabris CF, Hebert BM, Benjamin DJ, Beauchamp J, Cesarini D, et al. 2012. Most reported genetic associations with general intelligence are probably false positives. *Psychol. Sci.* 23:1314–23
10. Crow JF, Kimura M. 1970. *An Introduction to Population Genetics Theory*. New York: Harper & Row
11. Davies G, Tenesa A, Payton A, Yang J, Harris SE, et al. 2011. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol. Psychiatry* 16:996–1005
12. de Lange M, Snieder H, Ariens RA, Spector TD, Grant PJ. 2001. The genetics of haemostasis: a twin study. *Lancet* 357:101–5
13. de Moor MH, Costa PT, Terracciano A, Krueger RF, de Geus EJ, et al. 2010. Meta-analysis of genome-wide association studies for personality. *Mol. Psychiatry* 17:337–49

Vinkhuyzen et al.



14. Dudbridge F, Gusnanto A. 2008. Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* 32:227–34
15. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447:1087–93
16. Eisenhart C. 1947. The assumptions underlying the analysis of variance. *Biometrics* 3:1–21
17. Elston RC, Stewart J. 1971. A general model for the genetic analysis of pedigree data. *Hum. Hered.* 21:523–42
18. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, et al. 2012. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat. Genet.* 44:491–501
19. Evans DM, Frazer IH, Martin NG. 1999. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res.* 2:250–57
20. Falconer DF. 1965. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* 29:51–76
21. Falconer DF, Mackay TFC. 1996. *Introduction to Quantitative Genetics*. Essex: Longman
22. Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Soc. R. Soc. Edinb.* 52:399–433
23. Galton FRS. 1869. *Hereditary Genius. An Inquiry into its Laws and Consequences*. London: Macmillan
24. Galton FRS. 1886. Regression towards mediocrity in hereditary stature. *J. Anthropol. Inst. Great Br. Irel.* 15:17
25. Goddard M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136:245–57
26. Goddard ME, Meuwissen THE. 2005. The use of linkage disequilibrium to map quantitative trait loci. *Aust. J. Exp. Agr.* 45:837–45
27. Goldgar DE. 1990. Multipoint analysis of human quantitative genetic variation. *Am. J. Hum. Genet.* 47:957–67
28. Gottesman II. 1963. Heritability of personality: a demonstration. *Psychol. Monogr.* 77:1–21
29. Gregory AP, Dendrou CA, Attfield KE, Haghikia A, Xifara DK, et al. 2012. TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* 488:508–11
30. Guo SW. 1996. Variation in genetic identity among relatives. *Hum. Hered.* 46:61–70
31. Haddad S, Chen CA, Santangelo SL, Seddon JM. 2006. The genetics of age-related macular degeneration: a review of progress to date. *Surv. Ophthalmol.* 51:316–63
32. Haworth CM, Wright MJ, Luciano M, Martin NG, de Geus EJ, et al. 2010. The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Mol. Psychiatry* 15:1112–20
33. Henderson CR. 1950. Estimation of genetic parameters. *Ann. Math. Stat.* 21:2 (Abstr.)
34. Henderson CR. 1973. Sire evaluation and genetic trends. *J. Anim. Sci.* 1973:10–41
35. Hill WG, Goddard ME, Visscher PM. 2008. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* 4(2):e1000008
36. Hill WG, Weir BS. 2011. Variation in actual relationship as a consequence of mendelian sampling and linkage. *Genet. Res.* 93(1):47–64
37. Jia L, Landan G, Pomerantz M, Jaschek R, Herman P, et al. 2009. Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genet.* 5:e1000597
38. Johnson AM, Vernon PA, Feiler AR. 2008. Behavioral genetic studies of personality: an introduction and review of the results of 50+ years of research. In *Handbook of Personality Theory and Assessment*, ed. G Boyle, G Matthews, D Saklofske, pp. 145–73. London: Sage
39. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, et al. 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491:119–24
40. Keller MC, Medland SE, Duncan LE, Hatemi PK, Neale MC, et al. 2009. Modeling extended twin family data I: description of the Cascade model. *Twin Res. Hum. Genet.* 12:8–18
41. Kong X, Murphy K, Raj T, He C, White PS, Matise TC. 2004. A combined linkage-physical map of the human genome. *Am. J. Hum. Genet.* 75:1143–48
42. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467:832–38

43. Lander ES, Green P. 1987. Construction of multilocus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. USA* 84:2363–67
44. Lee SH, Harold D, Nyholt DR, Goddard ME, Zondervan KT, et al. 2013. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum. Mol. Genet.* 22:832–41
45. Lee SH, Wray NR, Goddard ME, Visscher PM. 2011. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* 88:294–305
46. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. 2012. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 28(19):2540–42
47. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, et al. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* In press
48. Lush JL, Lamoreux WF, Hazel LN. 1948. The heritability of resistance to death in the fowl. *Poult. Sci.* 27:375–88
49. Lynch M, Walsh B. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, Mass.: Sinauer Assoc.
50. Macgregor S, Cornes BK, Martin NG, Visscher PM. 2006. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Hum. Genet.* 120:571–80
51. Maes HH, Neale MC, Kendler KS, Martin NG, Heath AC, Eaves LJ. 2006. Genetic and cultural transmission of smoking initiation: an extended twin kinship model. *Behav. Genet.* 36:795–808
52. Maher B. 2008. Personal genomes: the case of the missing heritability. *Nature* 456:18–21
53. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–53
54. Martin NG, Eaves LJ. 1977. The genetical analysis of covariance structure. *Heredity* 38:79–95
55. McVean GA. 2002. A genealogical interpretation of linkage disequilibrium. *Genetics* 162:987–91
56. Meuwissen T, Goddard M. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics* 185:623–31
57. Orstavik KH, Magnus P, Reisner H, Berg K, Graham JB, Nance W. 1985. Factor VIII and factor IX in a twin population. Evidence for a major effect of ABO locus on factor VIII level. *Am. J. Hum. Genet.* 37:89–101
58. Patterson HD, Thompson R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–54
59. Pearson K, Lee A. 1903. On the laws of inheritance in man: I. Inheritance of physical characters. *Biometrika* 2:357–462
60. Poll H. 1914. Über zwillingsforschung als hilfsmittel menschlicher erbkunde. *Z. Ethnol.* 46:87–105
61. Posthuma D, de Geus EJ, Boomsma DI. 2001. Perceptual speed and IQ are associated through common genetic factors. *Behav. Genet.* 31:593–602
62. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38:904–9
63. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, et al. 2009. Cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.* 2:73–80
64. Rasmuson M. 1993. Variation in genetic identity within kinships. *Heredity* 70(3):266–68
65. Risch N, Lange K. 1979. Application of a recombination model in calculating the variance of sib pair genetic identity. *Ann. Hum. Genet.* 43:177–86
66. Ritland K. 1996. Marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* 50:1062–73
67. Schousboe K, Willemsen G, Kyvik KO, Mortensen J, Boomsma DI, et al. 2003. Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries. *Twin Res.* 6:409–21
68. Sham PC, Purcell S. 2001. Equivalence between Haseman-Elston and variance-components linkage analyses for sib pairs. *Am. J. Hum. Genet.* 68:1527–32
69. Siemens HW. 1924. Die zwillingspathologie. *Mol. Gen. Genet.* 35:311–12
70. Silventoinen K, Sarmalisto S, Perola M, Boomsma DI, Cornes BK, et al. 2003. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* 6:399–408

112 Vinkhuyzen et al.



71. Smith NL, Chen MH, Dehghan A, Strachan DP, Basu S, et al. 2010. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (cohorts for heart and aging research in genome epidemiology) consortium. *Circulation* 121:1382–92
72. Speed D, Hemani G, Johnson MR, Balding DJ. 2012. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* 91:1011–21
73. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* 42:937–48
74. Strandén I, Garrick DJ. 2009. Technical note: derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.* 92:2971–75
75. Tenesa A, Haley CS. 2013. The heritability of human disease: estimation, uses and abuses. *Nat. Rev. Genet.* 14:139–49
76. van der Harst P, Zhang W, Leach IM, Rendon A, Verweij N, et al. 2012. Seventy-five genetic loci influencing the human red blood cell. *Nature* 492:369–75
77. VanRaden PM. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–23
78. Vinkhuyzen AA, Pedersen NL, Yang J, Lee SH, Magnusson PK, et al. 2012. Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. *Transl. Psychiatry* 2:e102
79. Vinkhuyzen AAE, van der Sluis S, Maes HHM, Posthuma D. 2010. Reconsidering the heritability of IQ, taking into account assortative mating and cultural transmission. *Behav. Genet.* 42:187–98
80. Visscher PM. 2004. Power of the classical twin design revisited. *Twin Res. Off. J. Int. Soc. Twin Stud.* 7:505–12
81. Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90:7–24
82. Visscher PM, Hopper JL. 2001. Power of regression and maximum likelihood methods to map QTL from sib-pair and DZ twin data. *Ann. Hum. Genet.* 65:583–601
83. Visscher PM, Macgregor S, Benyamin B, Zhu G, Gordon S, et al. 2007. Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am. J. Hum. Genet.* 81:1104–10
84. Visscher PM, McEvoy B, Yang J. 2010. From Galton to GWAS: quantitative genetics of human height. *Genet. Res. (Camb.)* 92:371–79
85. Visscher PM, Medland SE, Ferreira MA, Morley KI, Zhu G, et al. 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* 2:e41
86. Visscher PM, Yang J, Goddard ME. 2010. A commentary on “common SNPs explain a large proportion of the heritability for human height” by Yang et al. (2010). *Twin Res. Hum. Genet.* 13:517–24
87. Wellcome Trust Case Control Consort. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–78
88. Whitfield JB, Martin NG. 1985. Genetic and environmental causes of variation in renal tubular handling of sodium and potassium: a twin study. *Genet. Epidemiol.* 2:17–27
89. Wray NR. 2005. Allele frequencies and the r^2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin Res. Hum. Genet.* 8:87–94
90. Wright S. 1920. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proc. Natl. Acad. Sci. USA* 6:320–32
91. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–69
92. Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88:76–82
93. Yang J, Lee T, Kim JE, Cho MC, Han BG, et al. 2013. Ubiquitous polygenicity of human complex traits: genome-wide analysis of 49 traits in Koreans. *PLoS Genet.* 9(3):e1003355
94. Yang J, Loos RJ, Powell JE, Medland SE, Speliotes EK, et al. 2012. FTO genotype is associated with phenotypic variability of body mass index. *Nature* 490:267–72
95. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, et al. 2011. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* 43:519–25
96. Yule, G. U. 1902. Mendel's laws and their probable relation to intra-racial heredity. *New Phytol.* 1:193–207; 222–38
97. Zuk O, Hechter E, Sunyaev SR, Lander ES. 2012. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. USA* 109:1193–98