

## PNAS

Contributed by Eric S. Lander, December 5, 2011 (sent for review October 9, 2011)

genome-wide association studies | statistical genetics

The prevailing view among human geneticists has been that the explanation for missing heritability lies in the numerator, that is, in additional variants remaining to be discovered. Much debate has focused on whether these additional variants are common alleles (frequency  $\geq 1\%$ ) with moderate-to-small effects or rare alleles

To avoid being fooled by phantom heritability, one might hope to be able to recognize when traits involve genetic interactions, for example, based on population data (such as phenotypic correlations among close relatives) or genetic data (such as pairwise tests

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1119675109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1119675109/-DCSupplemental).

of epistasis). We show, however, that this task may be difficult. For the case of Crohn's disease above, detecting the genetic interactions may require sample sizes in the range of 500,000.

In short, genetic interactions may greatly inflate the apparent heritability without being readily detectable by standard methods. Thus, current estimates of missing heritability are not meaningful, because they ignore genetic interactions.

Finally, we present a method to estimate  $h^2_{all}$  that is consistent not only for additive traits but for any genetic architecture. The method involves the study of isolated populations. It may provide a path forward for accurately measuring explained and missing heritability.

Extensive mathematical details and extensions are provided in *SI Appendix*. A Matlab software package used for the mathematical calculations is available at <http://www.broadinstitute.org/mpg/hc>.

## Results

**Quantitative and Disease Traits.** Quantitative traits are assumed to depend on genotype  $G$  and environment  $E$ , according to a function  $P = \Psi(G, E)$ . Here,  $G = (g_1, g_2, \dots, g_n)$  is the diploid genotype at  $n$  biallelic variant sites across the genome,  $g_i$  is the number of copies (0, 1, or 2) of a designated allele at the  $i$ th site, and  $f_i$  is the frequency of the designated allele. The variant sites are assumed to be in linkage equilibrium. The environment  $E$  may involve both a "shared" environment that is shared among pairs of relatives and a "unique" environment that is specific to each individual, which includes stochastic noise.

Disease traits are given by a binary function  $\Delta(G, E)$  that is assumed to arise from a liability threshold model. Specifically, there is an underlying (and unobserved) quantitative trait  $P = \Psi(G, E)$ , called a "liability." For a specified threshold  $\tau$ , individuals are affected ( $\Delta = 1$ ) if  $\Psi(G, E) \leq \tau$  and unaffected ( $\Delta = 0$ ) otherwise. (This condition is often equivalently defined as  $\Delta = 1$  if  $\Psi(G, E) \geq \tau$ .)

For convenience, we assume throughout that  $P$  has been normalized to have mean 0 and variance 1. With  $\text{Var}(P) = 1$ , the amount of variance explained by a factor is equal to the proportion of variance explained.

**Broad-Sense vs. Narrow-Sense Heritability.** Heritability is measured in two ways: broad-sense heritability  $H^2$  and narrow-sense heritability  $h^2$ .

Broad-sense heritability  $H^2$  measures the full contribution of genes. It is defined as  $H^2 = V_G/\text{Var}(P)$ , where  $V_G$  is the total variance due to genes. [Specifically,  $V_G = \text{Var}(P) - \text{Var}(P|G)$ , where  $\text{Var}(P|G)$  is the phenotypic variance between genetically identical individuals.]  $H^2$  is the relevant quantity for clinical risk assessment, because it measures our ultimate ability to predict phenotype from genotype.

By contrast, narrow-sense (or additive) heritability  $h^2$  is meant to capture the "additive" contribution of genes to the trait: It is the maximum variance that can be explained by a linear combination of the allele counts,  $g_i$ . Although  $h^2$  is a less intuitive concept, it is routinely used to measure progress toward explaining the genetic basis of a trait because one can readily calculate the contribution of individual loci to  $h^2$ , as described below.

**Explained, Missing, and Phantom Heritability.** We next define "explained" and "missing" heritability, focusing on narrow-sense heritability  $h^2$ . Let  $h^2_S$  (or  $h^2_{known}$ ) denote the proportion of the phenotypic variance explained by a set  $S$  of known variants, and  $h^2_{all}$  ( $=h^2$ ) denote the proportion of the phenotypic variance explained by all variants that affect the trait. For the variants in  $S$ , the proportion of "explained heritability" is  $\pi_{explained} = h^2_{known}/h^2_{all}$  and "missing heritability" is  $\pi_{missing} = 1 - \pi_{explained}$ . When all trait-associated variants have been found,  $\pi_{missing} = 0$ .

Human geneticists typically use a "bottom-up" approach to estimate the numerator and a "top-down" approach to estimate the denominator.

**Bottom-up.** The numerator  $h^2_{known}$  is straightforward to estimate, based on the effects of the individual variants. The variance explained by the  $i$ th variant is  $V_i = 2f_i(1 - f_i)\beta_i^2$ , where  $f_i$  is the frequency and  $\beta_i$  is the additive effect of the locus (defined as the regression coefficient of the phenotype  $P$  on the single-locus genotype  $g_i$ ). Under linkage equilibrium, the variance explained by a set  $S$  of variants is the sum over the individual loci:  $V_{known} = V_S = \sum_{i \in S} V_i$ . Because  $\text{Var}(P) = 1$ , we have  $h^2_{known} = V_{known}$  and  $h^2_{all} = V_{all}$ . We can thus estimate  $h^2_{known} = \sum_i 2f_i(1 - f_i)\beta_i^2$  from the allele frequencies and effect sizes estimated in a genome-wide association study.

**Top-down.** The problem comes in estimating the denominator  $h^2_{all}$ . Because not all variants are known, human geneticists must infer their total contributions indirectly, typically via a top-down quantity based on phenotypic correlations in a population. We refer to such quantities as "apparent heritability" and denote them by such symbols as  $h^2_{pop}$ .

Missing heritability is then estimated by assuming that  $h^2_{all} = h^2_{pop}$  and obtaining an estimate of  $h^2_{pop}$ . The problem is that  $h^2_{all}$  and  $h^2_{pop}$  are not guaranteed to be equal unless the trait is strictly additive, that is, involves neither gene-gene ( $G$ - $G$ ) nor gene-environment ( $G$ - $E$ ) interactions. For traits with genetic interactions,  $h^2_{pop}$  may significantly exceed  $h^2_{all}$ . If so, even when all variants have been discovered, the estimate of  $\pi_{missing}$  will not converge to zero with increasing sample size. Instead, it converges to  $1 - (h^2_{all}/h^2_{pop})$ , which we call the phantom heritability,  $\pi_{phantom}$ .

The term heritability and the symbol  $h^2$  are often used in the literature to refer to the true heritability  $h^2_{all}$  and to several definitions of apparent heritability  $h^2_{pop}$ , despite the fact that these various quantities need not be equal (*SI Appendix, section 1.5*). We have introduced distinct terminology and notation to avoid confusion about these important differences.

Analogous definitions can be made for broad-sense heritability  $H^2$ . In this case, it is easy to estimate the top-down quantity, but there is currently no practical way to estimate the bottom-up quantity (*SI Appendix, section 12*). As a result, human geneticists rarely attempt to estimate the proportion of the broad-sense heritability explained by a set of loci.

**Assuming Additivity.** We next describe the typical framework for analyzing human traits, noting why the equality  $h^2_{all} = h^2_{pop}$  depends on the assumption of additivity. We focus on one measure of apparent heritability,  $h^2_{pop}(ACE)$ , which considers additive genetic, common environmental and unique environmental variance components, but discuss alternative measures in *SI Appendix, section 1.3*.

**Quantitative traits.** A commonly used definition for apparent heritability is  $h^2_{pop}(ACE) = 2(r_{MZ} - r_{DZ})$ , where  $r_{MZ}$  and  $r_{DZ}$  are the phenotypic correlations between monozygotic twins and dizygotic twins, respectively (14). (The measure is based on the ACE model of twin studies.) One can show that

$$h^2_{pop}(ACE) = h^2_{all} + W, \text{ with} \quad [1]$$

$$W = \left[ \sum_{(i,j) \neq (1,0)} 2 \left( 1 - 2^{-(i+2j)} \right) V_{A^i D^j} \right] \geq 0,$$

where  $V_{A^i D^j}$  denotes the (nonnegative) variances due to all possible  $i$ th-order additive interactions and  $j$ th-order dominance interactions among loci (*SI Appendix, section 1*). The key point is that, if there are any genetic interactions, then  $W > 0$ , and  $h^2_{pop}(ACE)$  overestimates  $h^2_{all}$ . Unfortunately, there has been no way to estimate  $W$  from population data. In most human genetic studies, the "solution" has been simply to make the (usually unstated) assumption that there is no genetic interaction, that is, that  $W = 0$ . Typically, the studies assume a strictly additive model. (Some studies allow dominance terms at each locus, but they invariably assume additivity across loci, that is, no genetic interactions.)

Assuming a strictly additive model, the genetic architecture takes the form

$$\Psi(G, E) = \left[ \alpha + \sum_i \beta_i g_i \right] + \varepsilon, \quad [2]$$

with the two terms each being roughly normally distributed with mean 0 and with  $\Psi$  being normalized to have variance 1. Under this model, the variance of the first term is the narrow-sense heritability  $h^2_{all}$ . The environmental noise  $\varepsilon$  consists of shared and nonshared environments, with a vector  $c_R$  denoting the proportion of the environmental variance  $Var(\varepsilon) = 1 - h^2_{all}$  that is shared between relatives of type R. (For example,  $c_{sib}$  is the proportion of environment shared among sibs.) We will refer to this additive model as  $A(h^2, c_R)$ .

The additive model has many elegant properties. If  $\rho(R)$  denotes the phenotypic correlation between relatives of type R, then

$$\rho(R) = h^2 \gamma_R + (1 - h^2) c_R. \quad [3]$$

Here,  $\gamma_R$  is the genetic relatedness between relatives of type R ( $\gamma_R = 1, 1/2, 1/4$ , and  $1/8$  for MZ twins, sibs, grandparent-grandchild, and first cousins). The phenotypic correlation is proportional to genetic relatedness under the additive model with no shared environmental variance.

**Disease traits.** Disease traits are traditionally assumed to follow a liability threshold model, where the unseen liability  $\Psi$  follows the additive model above and disease occurs if  $\Psi \leq \tau$ . We will refer to this additive disease model as  $A_\Delta(h^2, c_R, \mu)$ , where  $\mu = Prob(\Psi \leq \tau)$  denotes the disease prevalence. The model parameters ( $h^2, c_R, \mu$ ) completely determine the epidemiologically observable quantities ( $\mu, \lambda_{MZ}, \lambda_{sib}, \dots$ ), where  $\mu$  is the disease prevalence and  $\lambda_R$  is the increased risk to relatives of type R (15).

To apply the model to a disease, one fits the model parameters based on observable quantities. (Geneticists often assume that  $c_R = 0$ , in which case the remaining two parameters can be fit based on  $\mu$  and  $\lambda_{MZ}$ .) For genetic variants associated with the disease, one then uses the model to convert an observed increase in disease risk to an inferred additive effect on the liability scale. Heritability calculations are performed not on the observed disease status but on the unseen liability scale. One advantage of using the liability scale is that heritability calculations tend to be robust to uncertainty about disease prevalence. (See *SI Appendix, section 2* for details, including the use of both  $\lambda_{MZ}$  and  $\lambda_{sib}$  to deal with shared environment.)

**Genetic Interactions Create Phantom Heritability.** What will happen if a geneticist analyzes a trait that involves genetic interactions under the erroneous assumption that it is additive? To explore this question, we introduce a simple and biologically plausible class of models.

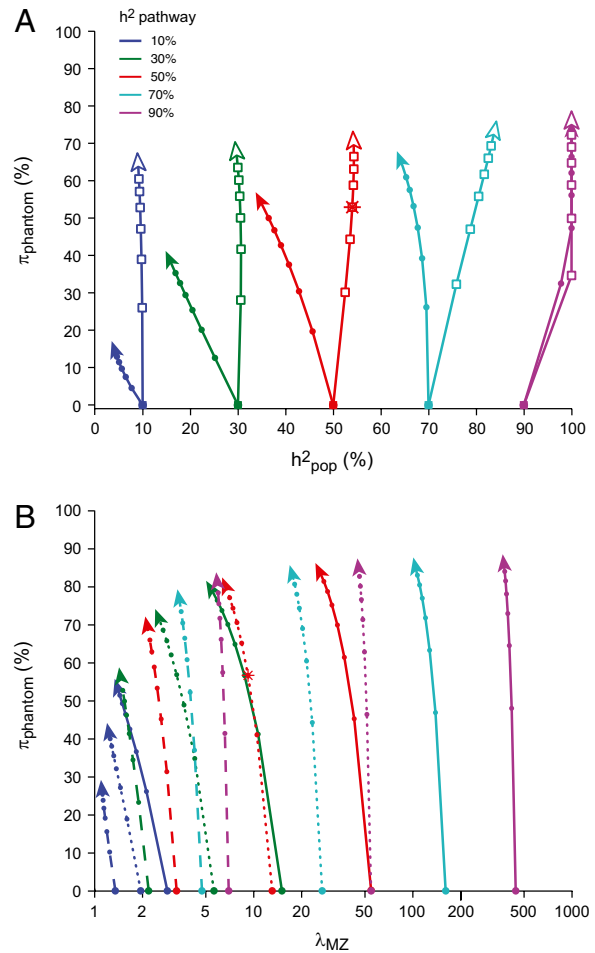
**Quantitative traits.** Biological processes often depend on the rate-limiting value among multiple inputs, such as the levels of components of a molecular complex required in stoichiometric ratios, reactants required in a biochemical pathway, or proteins required for transcription of a gene. We thus define a limiting pathway model, in which a trait  $P$  depends on the rate-limiting input from  $k \geq 1$  biological processes. For simplicity, we will assume that the inputs,  $\Psi_1, \Psi_2, \dots, \Psi_k$ , all follow the standard additive model in Eq. 2 above, each with exactly the same parameters,  $h^2_{pathway}$ , and  $c_R$ . Apart from the fact that the  $\Psi_i$  are roughly normal, we place no restrictions on the number or allele frequencies of the causal variants.

We define the trait  $LP(k, h^2_{pathway}, c_R)$  to be the minimum value of the  $\Psi_i$ . For a single pathway ( $k = 1$ ), the definition reduces to the simple additive model. What happens for  $k > 1$ ?

Let us consider a specific example:  $P^* = LP(4, 50\%, c_R)$ , with  $c_{sib} = 50\%$  (yielding shared environmental variance  $V_c = 27\%$ ) and  $c_R = 0$  for other relatives. Suppose that a geneticist analyzes  $P^*$  under the standard (but erroneous) assumption that it is

additive. Because we know the true genetic architecture (although the geneticist does not), we can calculate the exact value of all relevant parameters (*SI Appendix, section 3*). Because we are interested in asymptotic bias, we ignore sampling variation.

The geneticist would start by estimating the apparent heritability to be explained. The observed phenotypic correlation among twins is  $(r_{MZ}, r_{DZ}) = (62.4\%, 35.4\%)$ , yielding  $h^2_{pop} = 2(r_{MZ} - r_{DZ}) = 54.0\%$ . The geneticist would then conduct a genetic study, identify variants associated with the trait, estimate their effect sizes, estimate the heritability  $h^2_{known}$  explained by the variants, and compare it to the estimated value of  $h^2_{pop}$ . Assuming the sample is so large that all variants are identified (although the geneticist does not know this),  $h^2_{known}$  will be the true heritability,  $h^2_{all} = 25.4\%$ .



**Fig. 1.** Phantom heritability under the limiting pathway model. (A) Quantitative trait model  $LP(k, h^2_{pathway}, c_R)$ . For various parameters, curves show apparent heritability  $h^2_{pop}(ACE)$  and phantom heritability  $\pi_{phantom}$ . Curves connect points with various values of  $k$  (1, 2, 3, 4, 5, 6, 7, 10, with tip of arrow at  $k = 10$ ), and for specific values of  $h^2_{pathway}$  (10, 30, 50, 70, and 90%, indicated by color of the curve) and  $c_R$  (0% filled circles and arrows, 50% open boxes and arrows). The red asterisk indicates example  $P^*$  referred to in the text. (The quantity  $h^2_{pop}$  can exceed 100%, as seen for some models with high heritability and for some real traits. In such cases, we set  $h^2_{pop} = 100\%$ .) Raw data are in *SI Appendix, Table 6*. (B) Disease model  $LP_\Delta(k, h^2_{pathway}, 0\%, \mu)$ . For various parameters, curves show value of  $\lambda_{MZ}$  and phantom heritability  $\pi_{phantom}$ . Values of  $k$  and  $h^2_{pathway}$  are as in A. Values of prevalence  $\mu$  are 0.1% (solid), 1% (dotted), and 10% (dashed). The red asterisk indicates example  $\Delta^*$  referred to in the text. In both A and B, as  $k$  increases, the trait becomes more nonlinear; phantom heritability increases to 50% and beyond. Raw data are in *SI Appendix, Table 7*.



Even though all variants have been discovered, they will appear to explain only 47% ( $=25.4/54.0$ ) of the apparent heritability,  $h^2_{pop}$ . The remaining 53% is phantom heritability, which will never be explained by additional variants. It is the result of analyzing the data under an erroneous model.

Similar results are obtained for a wide range of parameters. Fig. 1A shows results for  $k = 1-10$ ,  $h^2_{pathway} = 10-90\%$ , and  $c_R = 0$  or 50%. The phantom heritability grows steadily with  $k$ . A mathematical theorem (16) implies that  $\pi_{phantom} \rightarrow 100\%$  as  $k$  grows (SI Appendix, section 3.4).

**Disease traits.** We can similarly define a limiting pathway model for disease traits by applying a threshold to the LP model for quantitative traits. Specifically, we define the disease trait  $LP_{\Delta}(k, h^2_{pathway}, c_R, \mu)$  as occurring if and only if  $LP(k, h^2_{pathway}, c_R) \leq \tau$ , with  $\mu$  denoting the disease prevalence. The case  $k = 1$  again reduces to the additive model. What happens for  $k > 1$ ?

Again, let us consider a specific case:  $\Delta^* = LP_{\Delta}(3, 50\%, c_R, 1\%)$ , with  $c_R = 0\%$  for all relatives. Based on the observed relative risks to MZ and DZ twins, a geneticist would calculate that  $h^2_{pop} = 49.0\%$ . However, an infinitely large genetic mapping study would yield  $h^2_{known} (=h^2_{all}) = 21.2\%$ . Even though all variants had been identified, they would appear to explain only  $43.2\% = (21.2/49.0)$  of the apparent heritability  $h^2_{pop}$ . The remaining 56.8% is phantom heritability. Similar results are obtained for a wide range of parameters. Fig. 1B shows results for  $k = 1-10$ ,  $h^2_{pathway} = 10-90\%$ , and  $c = 0.1, 1$ , and 10%.

**Epistasis Is Common.** The results show that mistakenly assuming that a trait is additive can seriously distort inferences about missing heritability. From a biological standpoint, there is no a priori reason to expect that traits should be additive. Biology is filled with non-linearity: The saturation of enzymes with substrate concentration and receptors with ligand concentration yields sigmoid response curves; cooperative binding of proteins gives rise to sharp transitions; the outputs of pathways are constrained by rate-limiting inputs; and genetic networks exhibit bistable states.

Genetic studies in model organisms have long identified specific instances of interacting genes (17). Important examples include synthetic traits (e.g., 18), which occur only when multiple loci or pathways are all disrupted. With the advent of genome-wide mapping in controlled genetic backgrounds in model organisms, studies have begun to reveal that epistasis is pervasive. In the yeast *Saccharomyces cerevisiae*, Brem et al. (19) analyzed as quantitative traits the levels of gene transcripts in segregants of a cross between two strains. For each transcript, they found the strongest quantitative trait locus (QTL) in the cross and then, conditional on the genotype at this locus, identified the strongest remaining QTL. In 67% of cases, these two QTLs demonstrated epistatic interactions. In bacteria, Khan et al. (20) and Chou et al. (21) have recently demonstrated clear epistasis among collections of five mutations that increase growth rate. In mouse and rat, Shao et al. (22) analyzed a panel of chromosome substitution strains, with each strain carrying a different chromosome from a donor strain on a common recipient genetic background. For dozens of quantitative traits, the sum of the effect attributable to the individual donor chromosomes far exceeds (median eightfold) the total effect of the donor genome, indicating strong epistasis. Although genetic interactions are hard to detect in humans (see below), several cases involving variants with large marginal effects have been recently reported in Hirschsprung's disease, ankylosing spondylitis, psoriasis, and type 1 diabetes (SI Appendix, section 7.1).

Several arguments are sometimes offered in support of the assumption of additivity (e.g., linearity of responses to selection). We discuss the flaws in such reasoning (SI Appendix, section 11).

#### Can We Detect Genetic Interactions by Comparisons Across Relatives?

Can a geneticist avoid being fooled by phantom heritability by detecting a priori that a trait involves genetic interactions, based

on the phenotypic correlations between close relatives? The task turns out to be difficult even if we restrict attention only to LP models.

**Phenotypic distribution.** The phenotypic distribution of a quantitative trait would not reveal the presence of genetic interactions. The distribution for  $LP(k)$  traits with modest values of  $k$  (say,  $k \leq 10$ ) is reasonably similar to the normal distribution in the additive model (SI Appendix, Fig. 1). Moreover, deviations from perfect normality are common in real traits and are typically resolved by applying a transformation to the distribution.

**Sib correlations.** Phenotypic correlations among sibs would not reveal that a trait involves genetic interactions. For quantitative traits, the correlations ( $r_{MZ}$ ,  $r_{DZ}$ ) for the LP models above are similar to those seen for real traits: They fit comfortably within the range of values recently reported by Hill et al. (23) for 86 traits (SI Appendix, section 5.1). For disease traits, the relative risks ( $\lambda_{MZ}$ ,  $\lambda_{sib}$ ) for various LP models similarly resemble those seen for real traits, for example, those reported for 15 actual diseases by Wray et al. (24) (SI Appendix, section 5.2).

**Correlations among extended relatives.** That sib correlations alone do not distinguish between additive and nonadditive LP models is not surprising: For either model, one can select parameters that largely fit the observed correlations. One might expand the analysis by considering additional relatives. For a trait with no shared environment, the phenotypic correlation between relatives should decrease linearly with genetic relatedness ( $\gamma_R$ ) if the trait is additive (by Eq. 3), but should be concave up if the trait involves genetic interactions. In theory, one could test for genetic interactions by fitting different genetic models to the curve of phenotypic correlations among relatives. In practice, it is difficult to draw strong conclusions from such analysis. First, such tests essentially depend on fitting a handful of values (e.g., correlations for individuals with  $\gamma_R = 1, 1/2, 1/4$ , and  $1/8$ ) with limited precision. Second, differences in the degree of shared environmental variance between relative types can substantially alter the shape of the curve (SI Appendix, section 6).

**Examples: Crohn's disease and schizophrenia.** The problem of discerning genetic architecture from a few parameters can be illustrated by considering alternative models for real diseases.

For Crohn's disease, current GWAS have identified 71 risk loci. Assuming the disease follows an additive model, these known loci explain  $h^2_{known} = 10.8\%$  of the total phenotypic variance, or  $\pi_{explained} = 21.5\%$  of the heritability (assuming  $h^2_{all} = h^2_{pop} = 50\%$ ). Alternatively, one can define an  $LP(3)$  model that is consistent with the prevalence and sib risks. Under this model, the phantom heritability is  $\pi_{phantom} = 62.8\%$ . Genetic interactions would account for  $80\% [=62.8/(1 - 0.215)]$  of the currently missing heritability. The known variants would account for  $\pi_{explained} = 57.5\% [=21.5/(1 - 0.628)]$  of the true heritability  $h^2_{all} = 18.6\%$  (SI Appendix, section 6).

For schizophrenia, Risch (15) presented recurrence risks for various relative types ( $\gamma_R = 1, 1/2, 1/4$ , and  $1/8$ ). We fit an additive model and an  $LP(2)$  model to the data (SI Appendix, section 6). Both models fit well, yet the former has no phantom heritability, whereas the latter has phantom heritability of 46%.

**Can We Detect Genetic Interactions from Pairwise Epistasis?** Even though it is difficult to detect genetic interactions a priori based on population data such as sib correlations, one might still hope to detect epistasis among variants a posteriori once they have been mapped. Indeed, geneticists have tested for pairwise epistasis between loci, but have found few significant signals. Should failure to detect pairwise epistasis allay our concerns about phantom heritability? Unfortunately, the answer is no.

The reason is that individual interaction effects are expected to be much smaller than linear effects, and the sample size required to detect an effect scales inversely with the square of the effect size. If  $n$  loci had equivalent effects, the sample size to

detect the  $n$  loci would thus scale with  $n^2$ , whereas the sample size to detect their  $\sim n^2$  interactions scales with  $n^4$ .

Consider the  $LP(3)$  disease model  $\Delta^*$  discussed above, with phantom heritability of 56.8%. Suppose that we consider two variants with frequency 20% that contribute to different pathways and increase risk by 1.3-fold (which is a large effect relative to those typically seen in GWAS). The sample size required to detect the variants is  $\sim 4,900$  (with 50% power and genome-wide significance level of  $\alpha = 5 \times 10^{-8}$  in a genome-wide association study with an equal number of cases and controls), whereas the sample size required to detect their pairwise interaction is roughly 450,000 (at 50% power and an appropriate significance level to account for multiple hypothesis testing). A researcher who studied 100,000 samples would likely discover all of the loci but would find little evidence of epistatic interactions. The researcher might conclude that the genetic architecture is additive, although the phantom heritability is actually  $>50\%$ . In short, the failure to detect epistasis does not rule out the presence of genetic interactions sufficient to cause substantial phantom heritability. (We discuss other ways to potentially detect epistasis in *SI Appendix, section 7.5*.)

**Consistent Top-Down Estimator of  $h^2_{all}$ .** What we need is a top-down estimator  $h^2_{all}$  that is consistent not simply for additive traits but for any genetic architecture. Traditional approaches fail because they focus on phenotypic correlations between close relatives; this creates two problems: (i) Extensive allele sharing between close relatives makes it difficult to disentangle the effects of genetic interactions; and (ii) differences of shared environment between different relative types make it difficult to disentangle the effects of environment.

We can eliminate these problems by studying nearly unrelated individuals in a population. Specifically, one can (i) identify pairs of individuals whose probability of allele sharing at the causal loci differs slightly from the population average, and (ii) measure how their phenotypic similarity depends on their genotypic similarity.

This goal can be accomplished by studying recent genetically isolated populations (such as Iceland, Finland, the Hutterites, or the Amish), in which one can use dense genotyping to reliably detect large segments shared identical-by-descent (IBD) between individuals (*SI Appendix, section 8*). We have the following theorem.

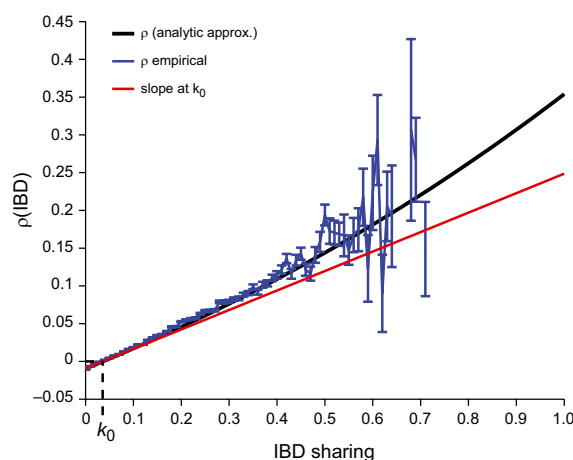
**Theorem 1.** Consider a population in which one can detect large segments shared IBD between individuals. Given two individuals  $I_i$  and  $I_j$ , let  $\kappa_{i,j} = \kappa(I_i, I_j)$  denote the proportion of their genomes shared in large IBD segments. Let  $\kappa_0$  denote the average value of  $\kappa$  across the pairs in the population.

Given a trait, let  $\rho(\kappa)$  denote the average phenotypic correlation between pairs of individuals who share proportion  $\kappa$  of their genomes in large IBD blocks. Regardless of the genetic architecture of the trait, the true heritability  $h^2_{all}$  equals  $h^2_{slope(\kappa_0)} = (1 - \kappa_0)\rho'(\kappa_0)$ , where  $\rho'(\kappa_0)$  is the rate of change of phenotypic correlation around the average sharing level of large IBD segments. Accordingly,  $h^2_{slope(\kappa_0)}$  provides a consistent top-down estimator for  $h^2_{all}$ .

The theorem applies to both quantitative traits and disease traits [with heritability measured on the disease (0,1) scale] with individuals sampled from the general population. The proof appears in *SI Appendix, section 8*, along with a version for individuals ascertained in a case-control study.

To apply this result in practice, one would (i) take a collection of individuals from the population; (ii) for each pair of individuals, calculate the product  $Q$  of the phenotype and the degree  $\kappa$  of IBD sharing; and (iii) estimate  $\rho'(\kappa_0)$  as the regression coefficient of  $Q$  on  $\kappa$ , for pairs with  $\kappa$  in a neighborhood around  $\kappa_0$ .

Fig. 2 illustrates the approach on simulated data for the quantitative trait  $P^*$  above, where  $h^2_{all} = 25.4\%$  and  $h^2_{pop} =$



**Fig. 2.** Estimating additive heritability from the slope of phenotypic correlation around mean IBD sharing. We performed simulations of the estimator in theorem 1. Genotype and phenotype data were generated for samples of 1,000 individuals chosen from an isolated population (mean IBD sharing  $\kappa_0 = 3.5\%$ , SD 5.7%) and the limiting pathway trait  $P^*$  (described in text) with 1,000 causal loci (see *SI Appendix, section 9* for details); results were averaged over 100 simulations. For each pair of individuals, we computed the product  $Z_1Z_2$  and the IBD sharing (*SI Appendix*). Blue error bars show mean and SD of expectation of  $Z_1Z_2$  for pairs in each 1% bin of IBD sharing, estimated from all such pairs across all 100 simulations. The black curve shows an analytic approximation for the mean phenotypic similarity  $r_R$  (*SI Appendix, section 3.3, Eq. 3.12*). The red line shows a least-squares linear regression line fitted using all pairs with IBD sharing in the interval  $[0, 2\kappa_0]$ . The average estimated slope (multiplied by  $1 - \kappa_0$ ) was  $0.258 \pm 0.082$ ; as expected from theorem 1, this is very close to the true heritability  $h^2_{all} = 0.254$  (and different from the apparent heritability,  $h^2_{pop} = 0.54$ ).

54%. With simulated data for 1,000 individuals with IBD sharing similar to that seen in Qatar (25), we estimate  $h^2_{slope(\kappa_0)} = 25.8 \pm 8.2\%$ , which is very close to the correct value of  $h^2_{all} = 25.4\%$ .

It is instructive to compare our approach with two elegant methods recently introduced by Visscher and colleagues, which inspired our own work. Both methods involve regressing phenotypic correlation on genotypic similarity. The first (26) measures genotypic similarity in terms of IBD within sib pairs—essentially measuring  $\rho'(1/2)$ , in our terminology. It eliminates the effects of shared environment by studying a single type of relative, but is confounded by genetic interactions because it studies close relatives (*SI Appendix, section 10*). The second (27) measures genotypic similarity in terms of identity by state across an SNP catalog for pairs of individuals in a population. As the authors note, the approach is not confounded by genetic interactions, but does not yield a consistent estimator because its sensitivity to causal variants falls with allele frequency. Nonetheless, this method yields a valuable lower bound on  $h^2_{all}$ .

## Discussion

The main points of this paper are that (i) current estimates of the proportion of heritability explained by known variants ( $\pi_{explained}$ ) implicitly assume that traits involve no genetic interactions; (ii) this assumption is not justified, because many models with interactions are equally consistent with available data; and (iii) under some of these models, the true value of  $\pi_{explained}$  may be much larger than current estimates. Accordingly, the widely held belief that missing heritability directly reflects the variance due to as-yet undiscovered variants is unjustified. Rather, missing heritability may be due in significant part to genetic interactions.

We focus here on a simple and biologically natural model, the limiting pathway model; it cannot readily be distinguished from an additive model based on population data or tests of pairwise

epistasis, yet entails substantial phantom heritability. Our focus on the LP model is not meant to imply that real traits necessarily follow this particular model; it simply provides an existence proof that erroneous assumptions may give rise to substantial missing heritability. We discuss more general multiple pathway models (*SI Appendix, section 4.4*), which also show substantial phantom heritability. (Beyond G–G interactions, we note that G–E interactions can produce additional phantom heritability.)

Importantly, we do not mean to propose that missing heritability is entirely, or even primarily, due to genetic interactions. On the contrary, many more causal variants are likely to exist, and to account for a significant part of the missing heritability. Discovery efforts should continue vigorously.

The case of Crohn's disease illustrates these points. The currently known loci can explain ~22%, ~58%, or more of the true heritability, depending on whether the disease follows an *LP*(1), *LP*(3), or other model. The available data cannot distinguish among the models. This spectacular degree of uncertainty undermines “inference by default,” for example, the frequent conclusion that rare variants must largely cause a disease, because common variants explain “too little” of the heritability. [Notably, a recent study of Crohn's disease (28) reported that the rare variants explained 10- to 20-fold less of the heritability than the common variants at 56 disease-associated loci.]

Given the dependence of results on genetic architecture, authors reporting proportions of heritability explained or missing should state clearly that the calculations are made under the arbitrary assumption that the trait is additive.

In LP models, phantom heritability increases with the number of pathways. More generally, traits with greater biological complexity may have greater phantom heritability. Current studies are broadly consistent with such a notion: The apparent heritability explained for “simpler” traits such as levels of fetal hemoglobin is greater than for “more complex” traits such as body-mass index or age at menarche (*SI Appendix, section 6.3*). Such differences may reflect both the number of loci and the genetic interactions underlying the traits.

The fraction of the apparent heritability of human traits due to genetic interactions cannot be inferred from available data,

although the pervasiveness of epistasis in experimental organisms suggests that the true heritability  $h^2$  of traits may be much lower than current estimates. (Lower values of  $h^2$  do not mean that traits are “less genetic” in the popular use of the term, which refers to the total contribution of genes,  $H^2$ . It simply means that additive effects comprise a smaller fraction of  $H^2$ .)

We describe a potential solution to overcome the problem of genetic interactions: Theorem 1 provides a top-down method to measure additive heritability that is consistent regardless of the underlying genetic architecture. In principle, the approach can provide an accurate assessment of heritability, as well as allow detection of the presence of genetic interactions by comparing top-down estimates obtained from different methods. To assess its practical utility, it will be necessary to apply it to appropriate data from isolated populations.

Finally, notwithstanding our focus here, we believe that concerns about missing heritability should not distract from the fundamental goals of medical genetics. Human genetic studies to discover variants associated with common traits should primarily be regarded as the analog to mutant hunts in model organisms, with the primary purpose being to identify the underlying pathways and processes. The key focus should be to study the biological role of the variants discovered so far. The proportion of phenotypic variance explained by a variant in the human population is a notoriously poor predictor of the importance of the gene for biology or medicine. [A classic example is the gene encoding HMGCoA reductase, which explains only a tiny fraction of the variance in cholesterol levels but is a powerful target for cholesterol-lowering drugs (1).] Ultimately, the most important goal for biomedical research is not explaining heritability—that is, predicting personalized patient risk—but understanding pathways underlying disease and using that knowledge to develop strategies for therapy and prevention.

**ACKNOWLEDGMENTS.** We thank David Altshuler, Jeffrey Barrett, Aravinda Chakravarti, Andrew Clark, David Golan, Peter Donnelly, Nick Patterson, Paz Polak, Alkes Price, David Reich, Peter Visscher, John Wakeley, and Noah Zaitlen for valuable discussions and comments. We thank Haley Hunter-Zinck and Andrew Clark for sharing data on inbreeding in Qatar. This work was supported in part by National Institutes of Health Grant HG003067 and by funds from the Broad Institute.

- Lander ES (2011) Initial impact of the sequencing of the human genome. *Nature* 470(7333):187–197.
- Manolio TA, Brooks LD, Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 118:1590–1605.
- Hirschhorn JN (2009) Genomewide association studies—Illuminating biologic pathways. *N Engl J Med* 360:1699–1701.
- Eichler EE, et al. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11:446–450.
- Manolio TA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753.
- Goldstein DB (2009) Common genetic variation and human traits. *N Engl J Med* 360:1696–1698.
- McClellan J, King MC (2010) Genetic heterogeneity in human disease. *Cell* 141:210–217.
- Slatkin M (2009) Epigenetic inheritance and the missing heritability problem. *Genetics* 182:845–850.
- Maher B (2008) Personal genomes: The case of the missing heritability. *Nature* 456(7218):18–21.
- Falconer DS, Mackay TF (1996) *Introduction to Quantitative Genetics* (Longman, Essex, UK), 4th Ed.
- Song YS, Wang F, Slatkin M (2010) General epistatic models of the risk of complex diseases. *Genetics* 186:1467–1473.
- Culverhouse R, Suarez BK, Lin J, Reich T (2002) A perspective on epistasis: Limits of models displaying no main effect. *Am J Hum Genet* 70:461–471.
- Franke A, et al. (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 42:1118–1125.
- Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, MA).
- Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 46:222–228.
- Berman SM (1964) Limit theorems for the maximum term in stationary sequences. *Ann Math Stat* 35:502–516.
- Carlberg O, Haley CS (2004) Epistasis: Too often neglected in complex trait studies? *Nat Rev Genet* 5:618–625.
- Ferguson EL, Horvitz HR (1989) The multivulva phenotype of certain *Caenorhabditis elegans* mutants results from defects in two functionally redundant pathways. *Genetics* 123(1):109–121.
- Brem RB, Storey JD, Whittle J, Kruglyak L (2005) Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* 436:701–703.
- Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332:1193–1196.
- Chou HH, Chiu HC, Delaney NF, Segrè D, Marx CJ (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332:1190–1192.
- Shao H, et al. (2008) Genetic architecture of complex traits: Large phenotypic effects and pervasive epistasis. *Proc Natl Acad Sci USA* 105:19910–19914.
- Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 4:e1000008.
- Wray NR, Yang J, Goddard ME, Visscher PM (2010) The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet* 6:e1000864.
- Hunter-Zinck H, et al. (2010) Population genetic structure of the people of Qatar. *Am J Hum Genet* 87(1):17–25.
- Visscher PM, et al. (2006) Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. *PLoS Genet* 2(3):e41.
- Yang J, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42:565–569.
- Rivas MA, et al.; National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC); United Kingdom Inflammatory Bowel Disease Genetics Consortium; International Inflammatory Bowel Disease Genetics Consortium (2011) Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 43:1066–1073.