

Assumptions and Properties of Limiting Pathway Models for Analysis of Epistasis in Complex Traits

Sven Stringer^{1,2,3*}, Eske M. Derks¹, René S. Kahn², William G. Hill⁴, Naomi R. Wray³

1 Department of Psychiatry, Amsterdam Medical Center, Amsterdam, The Netherlands, **2** Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands, **3** Queensland Brain Institute, University of Queensland, Brisbane, Australia, **4** Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom

Abstract

For most complex traits, results from genome-wide association studies show that the proportion of the phenotypic variance attributable to the additive effects of individual SNPs, that is, the heritability explained by the SNPs, is substantially less than the estimate of heritability obtained by standard methods using correlations between relatives. This difference has been called the “missing heritability”. One explanation is that heritability estimates from family (including twin) studies are biased upwards. Zuk et al. revisited overestimation of narrow sense heritability from twin studies as a result of confounding with non-additive genetic variance. They propose a limiting pathway (LP) model that generates significant epistatic variation and its simple parametrization provides a convenient way to explore implications of epistasis. They conclude that overestimation of narrow sense heritability from family data (‘phantom heritability’) may explain an important proportion of missing heritability. We show that for highly heritable quantitative traits large phantom heritability estimates from twin studies are possible only if a large contribution of common environment is assumed. The LP model is underpinned by strong assumptions that are unlikely to hold, including that all contributing pathways have the same mean and variance and are uncorrelated. Here, we relax the assumptions that underlie the LP model to be more biologically plausible. Together with theoretical, empirical, and pragmatic arguments we conclude that in outbred populations the contribution of additive genetic variance is likely to be much more important than the contribution of non-additive variance.

Citation: Stringer S, Derks EM, Kahn RS, Hill WG, Wray NR (2013) Assumptions and Properties of Limiting Pathway Models for Analysis of Epistasis in Complex Traits. *PLoS ONE* 8(7): e68913. doi:10.1371/journal.pone.0068913

Editor: Ali Torkamani, The Scripps Research Institute, United States of America

Received: September 25, 2012; **Accepted:** June 3, 2013; **Published:** July 30, 2013

Copyright: © 2013 Stringer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by grants to NRW from the Australian Research Council (FT0911360) and the Australian National Health and Medical Research Council (1011506) and a grant to EMD from the Netherlands Scientific Organization (NWO; project number 451-080-010). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: s.stringer@umcutrecht.nl

Introduction

A finding from genome-wide association studies for most complex traits is that the proportion of the phenotypic variance attributable to the additive effects of individual SNPs, i.e. the heritability explained by the SNPs, is substantially less than the estimate of heritability obtained from correlations of relatives using family data. Many explanations for this so-called ‘missing heritability’ have been proposed [1,2,3,4,5]. One explanation is that heritabilities from family (including twin) studies are overestimated. The problem of bias in heritability estimates has been much discussed in the quantitative genetic literature (e.g., [6,7,8,9]). For example, in the classical twin design of monozygotic (MZ) and dizygotic (DZ) twin pairs, there are only three essential statistics that can be estimated from their phenotypes, namely the MZ resemblance (such as covariance or correlation), the DZ resemblance, and the overall phenotypic variation in the sample. Therefore, only three variance components can be estimated, although many more genetic and non-genetic causal components of variance can be postulated to influence MZ and DZ resemblance. It is well recognized that estimates of heritability may be biased and that it is difficult to separate additive genetic from non-additive genetic components and to separate genetic from common (or shared) family environment components (e.g., [6,7,8,9]). Estimates of heritability

using phenotypic data from very distantly related individuals may have trivial bias from epistatic or common environment components compared to additive genetic components but are subject to very large sampling error. Human studies of distantly related individuals of sufficient size are simply not achievable. For disease traits, ascertainment bias in sampling of families for estimation of recurrence risks has long been recognized as a possible cause of inflated estimates of heritability [10,11]. Lastly, estimates of heritability for disease traits from twin cohorts (collected in restricted clinical settings) may be higher than those estimated from national cohort data, these differences most likely reflecting environmental factors including clinical practice [12].

Recently, Zuk et al. [13] revisited the overestimation of narrow sense (additive) heritability from family studies that could result from confounding with non-additive genetic variance. They referred to the difference between the expected value of the heritability estimated from family data and the ‘true’ heritability as ‘phantom heritability’. To illustrate their arguments, they proposed a limiting pathway (LP) model in which there are k pathway phenotypes, which are unobserved intermediate phenotypes. The phenotypically expressed trait value of an individual is the maximum of the individual pathway values. This model, they suggest, may be representative of biological processes that 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'. Under their LP model, each pathway phenotype includes only additive genetic effects but, for $k > 1$, non-additive genetic variance is generated for the expressed phenotype, and so the heritability of this phenotype is less than the expected estimate from a classical twin design analysis. As the magnitude of epistasis depends on k , the LP model provides a convenient way to explore the possible contribution of non-additive variation to missing heritability. They use this model to illustrate that over-estimation of heritability from pedigree data may explain an important proportion of missing heritability, but that quantifying this from available data is difficult. They advocate the continuation of association studies but argue that results should be reported acknowledging that heritabilities quoted from family studies may be overestimated. Nonetheless, their results may impact on the design of experiments seeking to identify disease or trait associated variants. It is therefore important to gauge carefully the likely relevance of their results.

Zuk et al. [13] consider their model to be simple and biologically natural. Their model is indeed simple and it usefully explores an epistatic model without needing to define genotypic effects at individual loci, because an infinitesimal model is assumed for each pathway. The pathways are assumed to be genetically independent and to have equal heritability, mean and variance. These are strong assumptions which may not be biologically plausible. For example, in human cells protein concentrations can be correlated and have different variances [14]. As complex traits are affected by many genes, individual genes will typically affect many complex traits [15,16]. Similarly, a single gene could affect multiple pathways, thereby creating a dependency between the affected pathways.

The purpose of this paper is twofold. First, we show that under the basic LP model, highly heritable quantitative traits produce phantom heritability only if the contribution of common variance is relatively large. Second, we explore the impact of the assumptions underlying the basic LP model. We extend the basic LP model to determine if their conclusions also hold after relaxing some assumptions to obtain a more biologically plausible model. Finally, we interpret the LP model in the context of other published studies.

Methods

Notation

Where convention allows, we use Greek symbols for population parameters and Roman for their estimates. In other cases we use a hat ($\hat{\cdot}$) notation to distinguish estimates from population parameters. Moreover, we use h^2 to represent the parameter of narrow sense heritability and h_{pop}^2 to represent the expected value of the heritability estimated from phenotypic data collected in the population. Here we consider estimation of h_{pop}^2 from twin data under the ACE (additive genetic, common environment, unique environment) model, which we denote with $h_{pop(ACE)}^2$. Similarly, we use c^2 and $c_{pop(ACE)}^2$ for the parameter and the expected value of its estimate of the proportion of variance attributable to the common environment under the ACE model.

The classical Twin Model

Like Zuk et al., we explore the LP genetic architecture through the variance components of the classical twin model. Under this model, only three independent parameters can be estimated from sets of MZ and DZ twins. One set of parameters is the phenotypic

variance σ_p^2 , the MZ correlation ρ_{MZ} and DZ correlation ρ_{DZ} . For MZ twins $\rho_{MZ} = (\sigma_G^2 + \sigma_C^2) / \sigma_p^2$ with σ_G^2 representing the total genetic variance and σ_C^2 the variance attributable to the common environment. The genetic variance can be broken down into additive σ_A^2 and non-additive components $\sigma_{NA}^2 = \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \dots$, the sum of all epistatic genetic variance components, with dominance represented by the subscript D. For DZ twins the phenotypic correlation is $\rho_{DZ} = (\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2 + \frac{1}{4}\sigma_{AA}^2 + \frac{1}{8}\sigma_{AD}^2 + \dots + \sigma_C^2) / \sigma_p^2$. As σ_C^2 is assumed to be equal in MZ and DZ twins $2(\rho_{MZ} - \rho_{DZ}) = (\sigma_A^2 + \frac{3}{2}\sigma_D^2 + \frac{3}{2}\sigma_{AA}^2 + \frac{7}{4}\sigma_{AD}^2 + \dots) / \sigma_p^2$ and $2\rho_{DZ} - \rho_{MZ} = (\sigma_C^2 - \frac{1}{2}\sigma_D^2 - \frac{1}{2}\sigma_{AA}^2 - \frac{3}{4}\sigma_{AD}^2 - \dots) / \sigma_p^2$. Usually heritability under the ACE model is estimated as $\hat{h}_{pop(ACE)}^2 = 2(r_{MZ} - r_{DZ})$ and the proportion of variance attributable to common environmental effects as $\hat{c}_{ACE}^2 = 2r_{DZ} - r_{MZ}$.

As well recognized [7,8,7,9], $\hat{h}_{pop(ACE)}^2$ is an upwardly biased estimate of the narrow sense heritability $h^2 = \sigma_A^2 / \sigma_p^2$. Likewise, when non-additive genetic variance is present $\hat{c}_{pop(ACE)}^2$ is a downwardly biased estimate of $c^2 = \sigma_C^2 / \sigma_p^2$, but if $2r_{DZ} > r_{MZ}$ we can conclude that common environment plays a role ($\hat{c}^2 > 0$). Also recognized in the quantitative genetics literature [7,6,9], and concluded by Zuk et al. [13], it is impossible to disentangle the contribution of epistasis and common variance based only on twin data. However, there are bounds on some parameters (and hence their estimates). From the equations provided above and by recognizing that variance components are non-negative, that variance components sum to σ_p^2 , and that MZ and DZ correlations are bounded between 0 and 1, some bounds are $\sigma_{NA}^2 / 2\sigma_p^2 \leq (\rho_{MZ} - 2\rho_{DZ} + \sigma_C^2 / \sigma_p^2) \leq \sigma_{NA}^2 / \sigma_p^2$, and $2\rho_{DZ} - \rho_{MZ} \leq \sigma_C^2 / \sigma_p^2$. In the absence of dominance and epistasis the lower bound of ρ_{DZ} is $\rho_{DZ} = \frac{1}{2}\rho_{MZ} + \frac{1}{2}\sigma_C^2 / \sigma_p^2$. We use these bounds to show in circumstances in which a large contribution from variance from epistasis is possible only if there is a large contribution to the variance from common environment. Whether this is plausible is trait dependent.

The basic LP Model

In the basic LP model for continuous traits [13], the final observed phenotype P is defined as the maximum (or equivalently the minimum) of k independent intermediate pathway phenotypes, $P_{max} = \max(P_1, \dots, P_k)$. The intermediate phenotypes P_i are completely additive, but the final phenotype P_{max} is not if $k > 1$. Zuk et al. [13] assumed an infinitesimal model for each pathway, so the basic LP model has three parameters: the number of (additive) genetic pathways k , the heritability of each pathway h_{path}^2 assumed to be constant across pathways, and the proportion of environmental variance $\sigma_{\epsilon_{path}}^2$ which is common among full siblings (including MZ and DZ twins) $c_{path} = \sigma_{C_{path}}^2 / \sigma_{\epsilon_{path}}^2$ (c_{sib} in [13]). For computational convenience the parameter c_{path} in the LP model is a proportion of the environmental variance to ensure a range between 0 and 1 independent of the value of h_{path}^2 . This should not be confused with the previously defined common variance c^2 which is proportional to the phenotypic variance.

When $k > 1$, the heritability h^2 of the expressed phenotype differs from the pathway heritability h_{path}^2 , likewise the proportion of environmental variance which is due to common environment in the pathway (c_{path}) is not necessarily equivalent to the analogous quantity at the final phenotype level (c). The basic LP model generates no dominance variance, but generates additive \times additive variance between loci from different pathways.

The Extended LP Model

In the basic LP model there are four important assumptions. All pathway phenotypes (i) have the same mean, (ii) the same variance (specifically $N(0,1)$), (iii) the same heritability, and (iv) are independent at the pathway level. As these assumptions are unlikely to be upheld in biological systems, we extend the basic LP model by relaxing some of them. In this extended LP model the means, variances and heritabilities may differ, and are defined by, respectively, $A\mu_{path}=(\mu_1,\dots,\mu_k)$, $\sigma_{path}^2=(\sigma_{P_1}^2,\dots,\sigma_{P_k}^2)$ and $h_{path}^2=(h_1^2,\dots,h_k^2)$. A general correlation matrix could be defined for the genetic relationship between pathways, but for simplicity we assume a uniform genetic correlation between all pathways, ρ_{path} when pathways are all positively correlated. Strong negative correlations between all pathways are not possible in general, so we consider the impact of negative correlations between pathways by dividing the pathways into two equally divided sets which are positively correlated by ρ within a set but negatively correlated by ρ_{path} between the sets.

The phenotype of pathway i (P_i) can be partitioned into additive genetic (A_i) and environmental (E_i) effects, $P_i=A_i+E_i$. No contribution of common environment is assumed between parents of the same child. The additive genetic variance in pathway i is $\sigma_{A_i}^2=\sigma_{P_i}^2 h_i^2$, and $\sigma_{E_i}^2=\sigma_{P_i}^2(1-h_i^2)$. Illustrating for $k=2$, the additive effects for both mothers and fathers are distributed as

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{A_1}^2 & \rho_{path}\sigma_{A_1}\sigma_{A_2} \\ \rho_{path}\sigma_{A_1}\sigma_{A_2} & \sigma_{A_2}^2 \end{pmatrix}\right)$$

The (unique) environmental (stochastic) effects for parents are assumed to be independent for each pathway and are distributed as

$$\begin{pmatrix} E_1 \\ E_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{E_1}^2 & 0 \\ 0 & \sigma_{E_2}^2 \end{pmatrix}\right)$$

As random mating is assumed, for offspring within a nuclear family the phenotype of pathway i for sibling j can be partitioned as

$$P_i(j)=A_i^{offspring}(j)+C_i+U_i(j)$$

since the additive genetic pathway values of an offspring are distributed with bivariate Mendelian sampling variance about the mean additive genetic values of their parents. C_i is the environmental effect of pathway i common to all siblings in a family, $C_i \sim \mathcal{N}(0,\sigma_{C_i}^2)$, and $U_i(j)$ is the environmental effect unique to sibling j , so for $k=2$,

$$\begin{pmatrix} A_1^{offspring}(j) \\ A_2^{offspring}(j) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \frac{1}{2}(A_1^{mother}+A_1^{father}) \\ \frac{1}{2}(A_2^{mother}+A_2^{father}) \end{pmatrix}, \begin{pmatrix} \frac{1}{2}\sigma_{A_1}^2 & \frac{1}{2}\rho_{path}\sigma_{A_1}\sigma_{A_2} \\ \frac{1}{2}\rho_{path}\sigma_{A_1}\sigma_{A_2} & \frac{1}{2}\sigma_{A_2}^2 \end{pmatrix}\right)$$

The proportion of environmental variance common for siblings at the pathway level is c_{path} (assumed to be the same for each pathway, and therefore the proportion of variance explained by common environment is $\sigma_{A_i}^2=\sigma_P^2 h_i^2$ hence $\sigma_{U_i}^2=(1-c_{path})\sigma_{E_i}^2$). Therefore, the extended LP model is a six parameter model: $LP_{ext}(k,A\mu_{path},\sigma_{path}^2,h_{path}^2,c_{path},\rho_{path})$. For example, the limiting pathway model $LP(k=2,h_{path}^2=0.8,c_{path}=0.5)$ could be modeled with the extended LP model as $LP_{ext}(k=2,A\mu_{path}=(0,0),\sigma_{path}^2=(1,1),h_{path}^2=(0.8,0.8),c_{path}=0.5,\rho_{path}=0)$. Zuk et al. [13] showed that the narrow-sense heritability of the observed phenotype in the population is $h^2=k\rho_{P_1,P}^2 h_{path}^2$, where $\rho_{P_1,P}$ is the correlation between the first pathway phenotype and the final phenotype. This definition assumes exchangeability and independence of the intermediate pathway phenotypes P_i . Under the extended LP model, the pathway phenotypes are non-exchangeable and correlated. As the additive model assumes $P=\mu+\sum_{i=1}^k \beta_i A_i + \varepsilon$, we can estimate the pathway coefficients $\hat{\beta}_i$ by regressing final phenotype P on the pathway genetic values A_i . The heritability estimate is a function of the regression coefficients, the additive values and the phenotype variance: $\hat{h}_{all}^2=Var(\sum_{i=1}^k \hat{\beta}_i \hat{A}_i)/\hat{\sigma}_P^2$. Unbiased estimates of h_{all}^2 and $h_{pop(ACE)}^2$ are reported as the mean of \hat{h}_{all}^2 and $\hat{h}_{pop(ACE)}^2$ across simulation replicates. As defined in Zuk et al. [13], the phantom heritability is $h_{phantom}^2=1-h_{all}^2/h_{pop(ACE)}^2$.

Simulation

For all simulations, we generated 50 independent samples of 100,000 families. Each family comprised two parents, an offspring, its MZ twin and its DZ twin. The phenotype of a parent for pathway i (P_i) was simulated as $P_i=A_i+E_i$. With random mating of parents, the phenotype of an offspring j for pathway i was simulated as

$$P_i(j)=A_i^{offspring}(j)+C_i+U_i(j)$$

with A_i , E_i , $A_i^{offspring}(j)$, C_i , and $U_i(j)$ drawn from their respective multivariate distributions. For monozygotic twins $A_i^{offspring}(j)=A_i^{offspring}(j')$. In all simulations, unless stated otherwise, the following parameters were used: $LP_{ext}(k=2,A\mu_{path}=(0,0),\sigma_{path}^2=(1,1),h_{path}^2=(0.8,0.8),c_{path}=0.5,\rho_{path}=0)$. From the final phenotypes of the offspring, the twin correlations r_{MZ} and r_{DZ} were calculated, resulting in a heritability based on the ACE model of $\hat{h}_{pop(ACE)}^2=2(r_{MZ}-r_{DZ})$ and phantom heritability $\hat{h}_{phantom}^2=1-\hat{h}_{pop(ACE)}^2/h_{pop(ACE)}^2$. The reported h^2 , c^2 , ρ_{MZ} , ρ_{DZ} , and $h_{phantom}^2$ are means across 50 simulation replicates and hence unbiased.

We first performed simulations to study the implications of the basic LP model with respect to common environmental effects. To explore bounds on variance components we simulated a range of basic LP models ($h_i^2 \in [0.1, 0.9], c_{path} = [0.1, 0.9], k \in \{1, 4\}$), and calculated h^2 and c^2 for each model. From each simulation we estimated ρ_{MZ} and ρ_{DZ} and plotted h^2 and c^2 as a function of ρ_{MZ} and ρ_{DZ} .

Subsequently we performed five simulations to study the effect of differences in (i) pathway mean, (ii) variance, and (iii) heritability on phantom heritability, and (iv & v) the effect of correlations between pathways on phantom heritability in the extended LP model. Simulations (i) to (iii) comprise a series of two-pathway models in each of which one parameter was changed: (i) the difference in pathway mean ($\mu_2 - \mu_1$). These values are in standard

deviation units since $\sigma_{P_i}^2 = 1$. (ii) The pathway variances differed, $\sigma_{P_1}^2 = 1$ and $\sigma_{P_2}^2$ was varied. (iii) The pathway heritabilities differed, $h_1^2 = 0.8$ and h_2^2 was varied. (iv) In this case a multiple pathway model was simulated with no common environment effects (i.e., $c_{path} = 0$), in which both the pathway correlations ρ_{path} and number of pathways k were varied. (v) As (iv), but with 10% of the variance of each pathway attributed to common environment (i.e., $c_{path} = 0.5$ when $h_{path}^2 = 0.8$).

Finally, to illustrate the extended LP model, we chose three continuous traits with different ACE-based heritability estimates from studies found in a recent twin research review paper [17]: (i) height in Danish male twins ($r_{MZ} = 0.89, r_{DZ} = 0.47, \hat{h}_{pop(ACE)}^2 = 0.84$) [18], (ii) triglyceride levels in blood in Swedish female twins from 20–29 years old ($r_{MZ} = 0.89, r_{DZ} = 0.47, \hat{h}_{pop(ACE)}^2 = 0.84$) [19], and (iii) high-fat dairy intake in UK male and female twins adjusted for age ($r_{MZ} = 0.23, r_{DZ} = 0.13, \hat{h}_{pop(ACE)}^2 = 0.20$) [20]. Based on these observed values and for different combinations of $k \in \{2, 4, 10\}$ and $\rho_{path} \in \{0, 0.2\}$, we report the estimates $\hat{h}_{phantom}^2, \hat{h}_{all}^2$, and $t \hat{c}^2$. As r_{MZ} and r_{DZ} are not model parameters but model outputs, the input parameters h_{path}^2 and c_{path} were chosen such that estimated r_{MZ} and r_{DZ} , based on the median of 50 simulations, reflected the observed values.

Results

Exploring Bounds of Variance Components

Although it is impossible to disentangle the contribution of non-additive genetic variance and common variance from twin data, there are some bounds on these parameters as illustrated in Figure 1, generated under the basic LP model, which is shown to yield more extreme non-additive genetic variance than many extended LP models. Each point in Figure 1 represents h_{all}^2 or c^2 as a function of ρ_{MZ}, ρ_{DZ} and k .

In comparisons of the left ($k = 1$) with the right panels ($k = 4$), each combination of ρ_{MZ} and ρ_{DZ} values is consistent with multiple basic LP models. In other words, the number of pathways cannot be derived from a pair of ρ_{MZ} and ρ_{DZ} values alone. As expected, for any ρ_{MZ} and ρ_{DZ} combination, the non-additive variance increases with the number of pathways, resulting in a lower narrow-sense heritability (h_{all}^2) estimate for an epistatic model ($k = 4$) compared to the additive model ($k = 1$) (panels A vs B). However, the contribution of environmental variance which is common for siblings (c) increases as well (panels C vs D). Therefore, as $2\rho_{DZ} - \rho_{MZ}$ increases for complex traits, important contributions from non-additive variance can be achieved only if accompanied by high c^2 . For example, if $\rho_{MZ} = 0.8$ and $\rho_{DZ} = 0.4$, an additive model ($k = 1; h^2 = 0.8$) implies no contribution of common variance ($c^2 = 0$), whereas a highly epistatic model (e.g., $k = 4; h^2 \in [0.2 - 0.4]$) is consistent only with $c^2 > 0.2$. More generally, for highly heritable traits ($h^2 > 0.6; k = 1$) a large amount of epistasis ($h^2 < 0.4; k = 4$) is consistent only with $c^2 > 0.2$. However, if $\frac{\rho_{MZ}}{2} > \rho_{DZ}$ then substantial phantom heritability need not be accompanied by large c^2 .

Properties of the Extended LP Model

Simulation results of the extended LP model are reported in Figure 2 for the effect of different parameters on the phantom heritability: (i) As the offset in mean between the two pathways increases, the phantom heritability decreases (panel A). One standard deviation difference in mean between pathway phenotypes (if $k = 2$) approximately halves the phantom heritability.

Clearly, as differences in offset become large, some pathways contribute little to the final phenotype, effectively decreasing the number of contributing pathways and hence the amount of epistasis. (ii) Differences in phenotypic variance between two pathways had no effect on phantom heritability (result not shown), because large phenotypic variance not only increases the probability of producing a maximum value, but also increases the probability of producing a minimum value. Across individuals both pathways contribute equally to the observed phenotype, but the mean and variance of the observed phenotype increases. Although the correlation between the final phenotype and the pathway phenotypes is higher for the pathway with the higher variance, the variance of the observed phenotype increases proportionally with the ratio of the two pathway variances, resulting in a constant heritability. (iii) As panel B shows, differences in heritability between pathway phenotypes have only a marginal effect on phantom heritability. (iv and v) Correlations between pathways affect the phantom heritability significantly (panels C and D). Positive correlations between pathways effectively limit the amount of epistasis, resulting in less phantom heritability. The larger the number of pathways, the larger the phantom heritability reduction (for any given pathway correlation). As the correlation between pathways approaches 1, the model approaches an additive single pathway model. This holds irrespective of the amount of common variance assumed, although common variance increases the phantom heritability slightly (panels D vs C). These results show that relaxing the assumptions of equal mean and uncorrelated pathways can substantially reduce the amount of phantom heritability. In contrast, negative correlations increase the amount of epistasis even if only one out of the k pathways is negatively correlated to the remaining positively correlated pathways, although the relative impact decreases as k increases (panels C and D).

Illustration for Three Traits

Table 1 shows the implications of the (extended) LP model for three continuous traits with increasing estimated heritability: high-fat dairy intake ($h_{pop(ACE)}^2 = 0.2$), triglyceride levels in blood ($h_{pop(ACE)}^2 = 0.54$), and height ($h_{pop(ACE)}^2 = 0.84$). The table illustrates two important points. First, assuming a larger amount of epistasis (i.e., larger k), not only implies increased phantom heritability and decreased narrow-sense heritability, but also implies a larger contribution of common variance. Second, assuming positive dependence between pathways ($\rho_{path} > 0$) reduces the amount of epistasis. In other words, increasing the number of pathways has less effect on phantom heritability, narrow-sense heritability, and the contribution of common variance, compared to a model with $\rho_{path} = 0$. Especially for traits with a large estimated ACE heritability, a high phantom heritability ($\hat{h}_{phantom}^2 > 0.61$) is only compatible with a scenario in which the percentage of common variance is high ($\hat{c}^2 > 0.39$).

Nonetheless, in some scenarios important phantom heritability is expected with negligible \hat{c}^2 , for example when $k = 2, \rho_{path} = 0$ we estimate $\hat{h}_{phantom}^2$ to be 0.25 and $\hat{c}^2 = 0.07$. We note that we selected examples with $\rho_{MZ} - 2\rho_{DZ} < 0$ to illustrate potential implications of the LP model. Hill et al. [21] reported an empirical distribution of $\rho_{MZ} - 2\rho_{DZ}$ distributed around zero, with interpretation that the distribution reflected sampling variance given the often small sample size. However, direct interpretation of the point estimates suggests that, since $\rho_{MZ} - 2\rho_{DZ} < 0$ in $\sim 50\%$ of cases, in these cases substantial phantom heritability could be

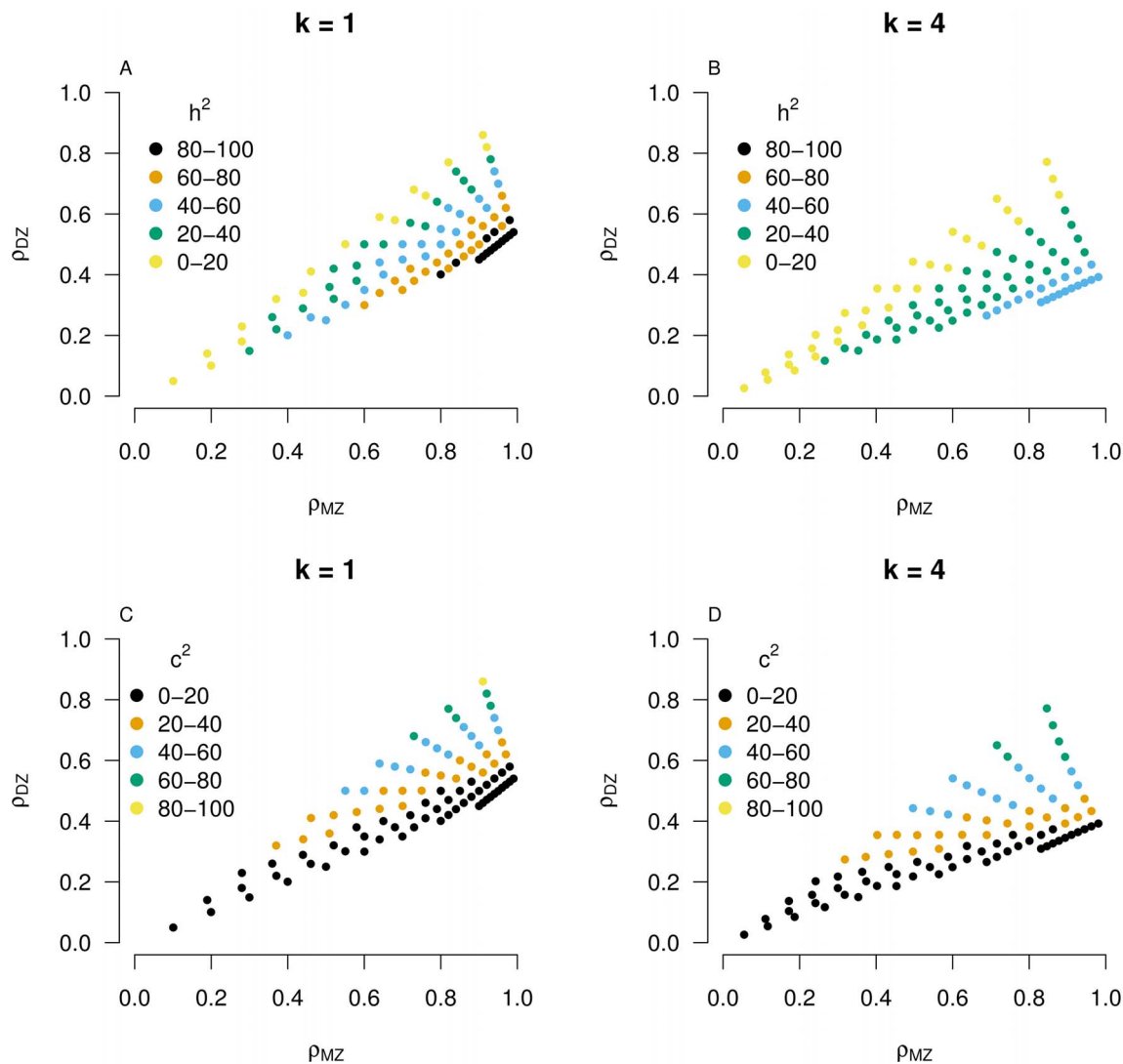


Figure 1. Mean narrow-sense heritability ($h^2 = \sigma_A^2 / \sigma_P^2$) (panels A and B) and proportion of phenotypic variance which is common for siblings ($c^2 = \sigma_C^2 / \sigma_P^2$) (panels C and D) across 50 simulation replicates color coded as function of MZ/DZ correlation and number of pathways (k) under the basic LP model. Given a combination of MZ/DZ correlations, a decrease in narrow-sense heritability (i.e., as k grows), implies an increase in contribution of common environment.
doi:10.1371/journal.pone.0068913.g001

present in the context of zero or weak common environmental variance.

Discussion

Bounds of Variance Components in Twin Studies

Using the basic LP model we explored constraints on combinations of parameters. For additive models (Figures 1A and 1C) all combinations of ρ_{MZ} and ρ_{DZ} lie in a region bounded by $\rho_{DZ} = \frac{1}{2}\rho_{MZ}$ (if $c=0$) and $\rho_{MZ} = \rho_{DZ}$ (if $c=1$). For epistatic models (Figures 1B and 1D) the bounds are evident from the wedge shape of permissible combinations of ρ_{MZ} and ρ_{DZ} in Figure 1. They show that when $\rho_{DZ} > \frac{1}{2}\rho_{MZ}$, substantial non-additive genetic variance can be accompanied only by unreasonably high σ_C^2 / σ_E^2 . This implies that, at least when $\rho_{DZ} > \frac{1}{2}\rho_{MZ}$ an underlying additive model is more plausible than a highly epistatic architecture. Specifically, as Table 1 illustrates, in highly heritable

traits with a small contribution of common variance, phantom heritability is likely to be small.

As noted by Zuk et al. [13] in their supplementary information, the amount of phantom heritability estimated depends on the method of estimation of h_{pop}^2 . The expected heritability estimate from regression of offspring phenotype on mid-parental phenotype ($h_{pop(PO)}^2$) is less than $h_{pop(ACE)}^2$ under the $k > 1$ basic LP model. Other factors could also contribute to differences between $h_{pop(ACE)}^2$ and $h_{pop(PO)}^2$ such as dominance and greater common environment of sibling compared to filial relations. Despite this, empirical observation ([7] pp. 172–173) does not, in general, suggest large differences between $h_{pop(ACE)}^2$ and $h_{pop(PO)}^2$, which is not consistent with an important role for phantom heritability (although sampling variation about estimates make it difficult to draw strong conclusions). Deconfounding of genetic and common environmental variance is possible, for the most part, by use of adopted away relatives. Very different estimates of correlations

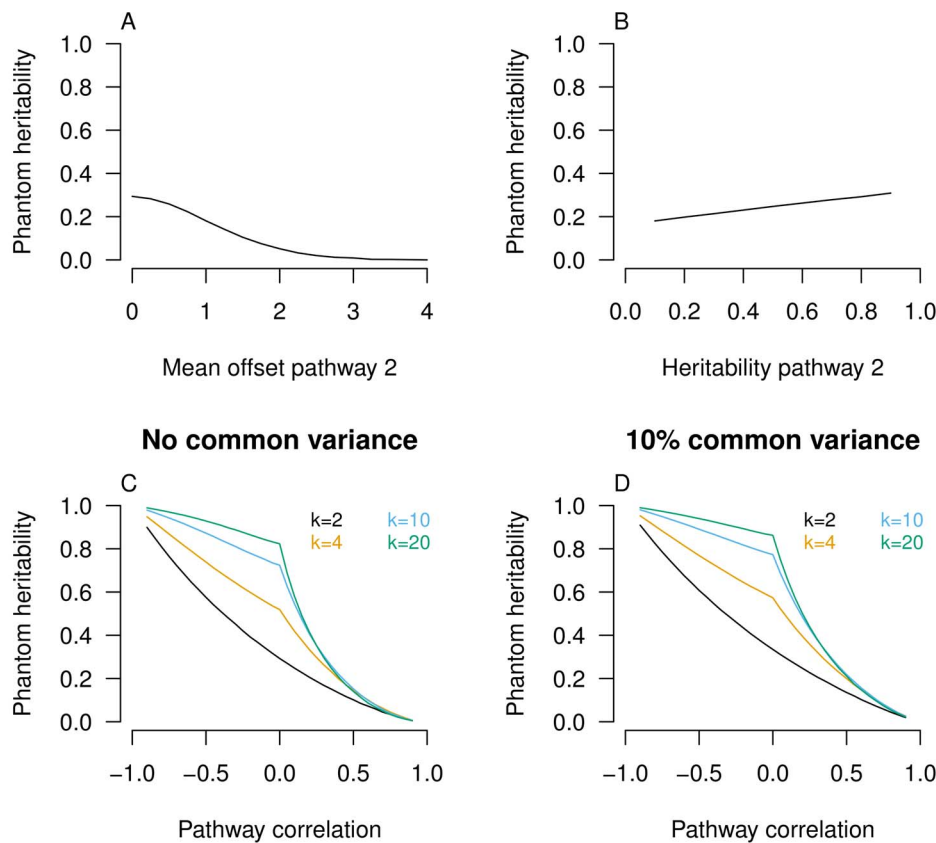


Figure 2. Phantom heritability under the extended LP model as a function of (A) differences in mean (sd unit) of two pathway phenotypes, $(\mu_2 - \mu_1)$ (B) changes in $h_{path_2}^2$, while $h_{path_1}^2 = 0.8$, (C) pathway correlations for different number of pathways, and (D) pathway correlations for different numbers of pathways assuming a total contribution of common environment of 10% for each pathway phenotype. Unless stated otherwise, number of pathways $k=2$, contribution of common family environment is zero (i.e., $c_{path}=0$), pathway heritability $h_{path}^2=0.8$, pathway correlations $\rho_{path}=0$.
doi:10.1371/journal.pone.0068913.g002

between adopted away siblings and those raised together is expected if phantom heritability is important, but adoption studies tend to support genetic estimates from twin studies [22].

The Extended LP Model

Zuk et al. proposed a simple and elegant model that allows exploration of the impact of epistasis on estimates of heritability without needing to define epistasis between individual loci. In fact, the basic LP model is a special case of the optimum pathway model proposed by Sewall Wright in 1935 [23], in which the expressed phenotype is the pathway value closest to a defined optimum, which could be, for example, the mean or median, rather than the maximum. These models include the additive model as a special case, but produce different amounts of epistasis as the number of pathways increases. Indeed any non-linear transformation of an additive genetic model, even the infinitesimal model, leads to non-additive variation; but Zuk et al. show that the basic LP model generates a phenotypic distribution close to normal, particularly when k is small. Under the basic LP model all pathway phenotypes have the same distribution and pathway heritabilities and pathways are uncorrelated. Biologically, these are very strong assumptions, not least since they invoke the infinitesimal model that implies independent contributions from many genomic sites in each pathway, and so we extended the basic LP model to allow correlated pathway phenotypes with different distributions and pathway heritabilities. Phantom heritability was

little affected by differences in variance and heritability between pathways. However, differences in mean phenotype and the presence of positive correlations between pathways can decrease the phantom heritability considerably, and negative correlations increase it. Our results show how the predicted importance of phantom heritability depends on implicit model assumptions, such that the problem of phantom heritability could be overstated. Drawing inferences about epistasis from the LP model.

There is much debate about the relative importance of non-additive versus additive genetic variance [24,25,26,21,27], summarized by Crow [28]. Central to the debate is that mutational studies demonstrate the ubiquity of epistasis in the classical sense, because genes interact in hierarchical systems to generate biological function [26]. However, in quantitative genetics it is the residual variation segregating in populations that determines differences amongst individuals not overall biological function [26]. Fisher suggested that epistasis was not important because usually there would be some scale transformation of phenotypic values to generate additive effects [26]. Indeed, this is the basis of models of complex disease where non-additivity on the observed scale can be transformed to an underlying additive scale. Furthermore, under mutation drift (neutral) models a high proportion of genetic variants are at frequencies near 0 or 1, so the presence of substantial epistatic interactions at the level of gene effects does not in itself generate appreciable epistatic variance, and contributions from epistatic interactions are detected as

Table 1. Phantom heritability ($\hat{h}_{phantom}^2$), narrow-sense heritability (\hat{h}^2) and percentage of common variance (\hat{c}^2) for three traits assuming varying number of pathways (k) and pathway correlations (ρ_{path}).

Trait	r_{MZ}	r_{DZ}	h_{ACE}^2	k	ρ_{path}	$\hat{h}_{phantom}^2$	\hat{h}^2	\hat{c}^2
High-fat dairy intake	0.23	0.13	0.20	2	0	0.15	0.17	0.04
				4	0	0.30	0.14	0.06
				10	0	0.56	0.09	0.10
				2	0.2	0.05	0.19	0.03
				4	0.2	0.13	0.17	0.04
				10	0.2	0.22	0.16	0.06
Triglyceride levels in blood	0.55	0.28	0.54	2	0	0.25	0.41	0.07
				4	0	0.49	0.28	0.16
				10	0	0.74	0.14	0.29
				2	0.2	0.16	0.45	0.05
				4	0.2	0.28	0.39	0.09
				10	0.2	0.38	0.33	0.15
Height	0.89	0.47	0.84	2	0	0.35	0.55	0.20
				4	0	0.61	0.33	0.39
				10	0	0.82	0.15	0.61
				2	0.2	0.26	0.62	0.17
				4	0.2	0.43	0.48	0.30
				10	0.2	0.56	0.37	0.43

Illustrated for observed values of the estimated heritability h_{ACE}^2 and underlying r_{MZ} and r_{DZ} , assuming a larger number of pathways implies higher phantom heritability, lower narrow-sense heritability, but also a larger contribution of common variance. Higher pathway correlations reduce these effects.
doi:10.1371/journal.pone.0068913.t001

additive variance [28,21]. These arguments are further strengthened under models that consider selection against mutations deleterious for fitness with pleiotropic effects on quantitative traits [29] as the proportion of variants with frequencies near 0 or 1 is even higher. Zuk et al. [13] incorrectly state in their Supplementary Information that the derivation in Hill et al. [21] applies only to pairs of loci, whereas in fact these were used for illustration, and the argument holds for multilocus epistasis. The elegance of the LP model of Zuk et al. [13] is that it is parametrized in terms of variances and so does not depend on the allele frequency distribution. Zuk et al. [13] (supplement page 45) counter Hill et al.'s analysis by arguing that most genetic variants contributing to complex traits cannot be at extreme frequencies because these would generate little variance. They illustrate with a two-locus example (their Supplementary Figure 9), but it shows a steep increase in total genetic variance from minor allele frequency of 0 to 0.1, nearing its maximum for minor allele frequency 0.1, where additive \times additive variance accounts for only 8% of the genetic variance. Furthermore, because the distribution of heterozygosity is approximately uniform over 0 to 1 under the neutral mutation drift model, all frequencies are expected to contribute approximately equally to the variance under an additive model. Empirical results also suggest that epistasis can generate little epistatic variance. For example, although many substantial epistatic effects have been detected for bristle number in *Drosophila* [30], bristle number expresses mostly additive variance in populations [16]. At face value these results may seem to be contradicted by recent results of the *Drosophila* Genetics Research Panel (DGRP) entitled "Epistasis dominates the genetic architecture of *Drosophila* quantitative traits" [31]. They reported data are from a GWAS undertaken on the 168 DGRP lines [32] and on gene frequency

differences between pools of lines scoring high and low for phenotypes following an advanced intercross (70 generations) from 40 of the DGRP lines [31]. They found no overlap of SNP associated effects between the two analyses, which they interpreted as presence of epistasis. However, the limited number of DGRP lines are underpowered for association analysis and show long range LD so effects of distantly located QTL are confounded and are less likely to match those found in the intercross study. Also, as the authors [31] note: "In fact, variation induced by all of the epistatic interactions identified in the present study could be largely explained by the marginal additive effects at the trait-associated loci".

Limiting Pathways in Context

The LP model was justified (Zuk et al., p1193 [13]) without reference as: "Here we show that simple and plausible models can give rise to substantial phantom heritability. 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." For biochemical pathways, at least, metabolic control theory has shown that 'rate limiting steps' are not a relevant concept, for rate of flux is a continuous function of activities at multiple stages of the pathway [33]. In a recent review Suarez and Moyes stated "The days have long passed when it was simply assumed that enzymes possessing allosteric regulatory properties were 'rate-limiting' [34]. It is now recognized that control of pathway flux is often distributed among many enzymes." And Fell's well-cited review [35] concludes "whatever criticisms might be made about any one of the experimental studies, it is significant that none have

provided support for the existence of unique ‘rate-limiting’ enzymes in pathways.” In quantitative genetic analysis of models of such pathways, it has been shown that a substantial proportion of the variance is additive [36].

The LP model was proposed to explain “missing heritability” in complex traits. Methods are now available to estimate variance attributable to all common genotyped SNPs rather than those identified as significant [37,38]. Simulations conducted under the LP model demonstrate that estimates of additive variance attributable to SNPs calculated using GCTA [39] are unbiased and not inflated by epistasis [40]. Applications of these methods to real data show that at least 40% of heritability estimated from family studies remains unexplained [41]. The number of associated common variants detected has increased with sample size [41] e.g., from 9 to 140 for Crohn’s Disease as case sample size increased from 2000 [42] to >2000[43]. The implication is that, to date, studies have been underpowered to detect common variants of realistic effect sizes, but that many exist, given that rare variants are much more prevalent, and that a very large number of rare variants also contribute exist which individually explain little variance but their cumulative contribution may be important. Collection of empirical data to test an additive only model is unlikely to be achievable in humans. In yeast, an elegant study designed to explore contributions of variance from different sources found substantial epistatic variance (median of 30%) for some of the 46 traits studied [44]. However, its relevance to human populations is limited, since all gene frequencies were one half (two-way cross design), conditions under which epistasis is likely to be maximized [21,28]. More relevant insight may be gained from outbred species. For example, in dairy cattle heritability and SNP associated effects are estimated from large numbers of half-sib daughters born, raised and milked at different farms. Therefore, their estimates are unlikely to be confounded with non-additive genetic or shared environmental effects [45]. For milk yield 79% (s.e. 5%) of the additive genetic variance is captured by SNPs [46]. That there is so little missing heritability can be explained by the smaller effective population size leading to longer linkage disequilibrium (LD) blocks than in humans and hence even rare alleles can be predicted by multiple SNPs. Traits that could reasonably be assumed to be under strong natural selection (so that very rare variants play an important role), such as fertility, have lower heritability (40%) and greater missing heritability (55% explained by common SNPs fitted together) [46]. The simplest explanation of why not all variance is explained by the SNPs is that even in livestock some causal variants are rare and in low LD with the SNPs. These results provide evidence that (when h_{pop}^2 is estimated accurately) additive effects can explain the majority of observed variance in a complex trait in an outbred population.

Disease Traits

Zuk et al. [13] expressed phantom heritability as $h_{phantom}^2 = 1 - \frac{h_{all}^2}{h_{pop}^2}$. For quantitative traits $h_{all}^2 = h_{P_{max}}^2$, where $h_{P_{max}}^2$ is the proportion of variance in P_{max} attributable to additive genetic factors. For disease traits they considered a liability threshold model, but did not assume disease to occur when the liability phenotype P_{max} exceeds the threshold truncated by the proportion 2μ , but instead defined disease to occur when a pathway phenotype exceeds the threshold truncated by the proportion $2\mu/k$, generating a total proportion, 2μ , of affected individuals when summed over all k pathways. This definition implies additional non-additive genetic variance, i.e., $h_{all}^2 < h_{P_{max}}^2$.

For example, using the 3-pathway model for Crohn’s Disease [13] with $h_{path}^2 = 0.476$ and $c_{path} = 0.16$ generates $h_{P_{max}}^2 = 0.283$, but $h_{all}^2 = 0.186$ for $2\mu = 0.001$. Under the extended LP model we showed for quantitative traits that results for a multiple pathway model converged to a single pathway model for positively correlated pathways; qualitatively this result also holds for disease traits.

The LP model was justified by Zuk et al as limiting pathways in a biological and biochemical sense. However, the same methodological approach could represent a heterogeneity model, generating a different interpretation of results. Under the LP model the final phenotype is considered the “true” phenotype and the non-additive genetic variance $h_{pop}^2 - h_{all}^2$ is real. In contrast, under a heterogeneity model, the pathways are the true phenotypes but inadequacies in phenotyping cause an inability to distinguish between biologically different classes of the observed disease. Hence, under a heterogeneity model, the measurable additive genetic variance h_{all}^2 may be much less than true additive variance of each subtype, but mostly $h_{pop}^2 - h_{all}^2$ could be viewed as “phantom non-additive genetic variance”, since the non-additive genetic variance results only from incorrectly treating multiple phenotypes as a single trait. In common complex genetic disease there have been notable advances in separation of diseases that originally were considered a single diagnostic class, e.g., diabetes, rheumatoid arthritis, breast cancer. Dilution of allelic effect size is a consequence of phenotypic heterogeneity in genetic association studies. For example, differentiation of breast cancer into ER-positive and ER-negative cancers has identified associated loci not possible from combining the case cohorts [47]. In psychiatric nosology it has long been recognized that diagnostic classes are likely to overarch heterogeneous etiology, recently explored in light of results from genomic studies [48]. Indeed, one motivation of genomic studies is to allow genetically informed nosology.

Conclusion

The results of Zuk et al. [13] provide a timely reminder of the well-recognized limitations of analyses based on twin and family data, which are often underpowered to separate additive genetic from common environmental effects [49] and non-additive effects. The (extended) LP model provides a useful framework to explore the possible contribution of non-additive genetic variance to complex traits. An important role for non-additive genetic action is attractive because gene interactions are ubiquitous at the functional level, yet this does not necessarily translate to important epistatic variance over and above variance detected as additive effects. For disease traits, empirical data can only be explained by non-additivity on the disease scale, but such non-additivity can be explained by scale transformations without needing to invoke more complex models. Using the framework of the extended LP model, and together with theoretical, empirical, and pragmatic arguments we conclude that although contributions from non-additive variance may be commonplace in complex traits, the contribution of additive genetic variance is likely to be much more important than that of non-additive variance. Ultimately, only empirical results can provide a satisfactory conclusion to the debate of missing heritability, but these may be elusive. Larger sample sizes should afford the power to identify common variants of smaller effect size and two-locus interactions. However, the heavy penalty of multiple testing will not allow exploration of higher order epistatic interactions implied by the LP model. Likewise, large sample sizes are unlikely to identify rare causal variants of small effect, since rare variants are likely to be population specific and large sample sizes from homogenous

ethnic groups simply may not exist. Zuk et al. [13] suggest a methodology for estimation of h_{all}^2 , but the required cohorts (large and from isolated populations) are also difficult to achieve. For disease traits the most tractable approach may be collection of large, informatively phenotyped cohorts to provide the building blocks that may allow clustering of cases based of combinations of genetic risk variants to be mapped onto phenotypic heterogeneity.

References

1. Maher B (2008) The case of the missing heritability. *Nature* 456: 18–21.
2. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747–753.
3. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, et al. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11: 446–450.
4. Clarke AJ, Cooper DN (2010) GWAS: heritability missing in action? *Eur J Hum Genet* 18: 859–861.
5. Stringer S, Wray NR, Kahn RS, Derks EM (2011) Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. *PLoS ONE* 6: e27964.
6. Falconer DS, Mackay TFC (1996) *Introduction to quantitative genetics*, 4th edition. Harlow, Essex: Longmans Green.
7. Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.
8. Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era - concepts and misconceptions. *Nat Rev Genet* 9: 255–266.
9. Sham P (1998) *Statistics in Human Genetics*. London: Wiley.
10. Haenszel W (1959) Some problems in the estimation of familial risks of disease. *J Natl Cancer Inst* 23: 487–505.
11. Guo SW (1998) Inflation of sibling recurrence-risk ratio, due to ascertainment bias and/or overreporting. *Am J Hum Genet* 63: 252–258.
12. Wray NR, Gottesman II (2012) Using summary data from the Danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. *Front Genet* 3: 1–12.
13. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 109: 1193–1198.
14. Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, et al. (2006) Variability and memory of protein levels in human cells. *Nature* 444: 643–646.
15. Hill WG, Zhang XS (2012) On the pleiotropic structure of the genotype-phenotype map and the evolvability of complex organisms. *Genetics* 190: 1131–1137.
16. Mackay TFC (2010) Mutations and quantitative genetic variation: lessons from *Drosophila*. *Philos T Roy Soc B* 365: 1229–1239.
17. van Dongen J, Slagboom PE, Draisma HHM, Martin NG, Boomsma DI (2012) The continuing value of twin studies in the omics era. *Nat Rev Genet* 13: 640–653.
18. Silventoinen K (2003) Determinants of variation in adult body height. *J Biosoc Sci* 35: 263–285.
19. Rahman I, Bennet AM, Pedersen NL, de Faire U, Svensson P, et al. (2009) Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. *Twin Res Hum Genet* 12: 286–294.
20. Teucher B, Skinner J, Skidmore PML, Cassidy A, Fairweather-Tait SJ, et al. (2007) Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Res Hum Genet* 10: 734–748.
21. Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 4: e1000008.
22. Petersen L, Sørensen TIA (2011) The Danish adoption register. *Scand J Public Health* 39: 83–86.
23. Wright S (1935) The analysis of variance and the correlations between relatives with respect to deviations from an optimum. *J Genet* 30: 243–256.
24. Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nat Rev Genet* 3: 11–21.
25. Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? *Nat Rev Genet* 5: 618–625.
26. Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9: 855–867.
27. Hemani G, Knott S, Haley C (2013) An evolutionary perspective on epistasis and the missing heritability. *PLoS Genet* 9: e1003295.
28. Crow JF (2008) Maintaining evolvability. *J Genet* 87: 349–353.
29. Eyre-Walker A (2010) Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc Natl Acad Sci U S A* 107: 1752–1756.
30. Mackay TF, Lyman RF (2005) *Drosophila* bristles and the nature of quantitative genetic variation. *Philos Trans R Soc Lond B Biol Sci* 360: 1513–1527.
31. Huang W, Richards S, Carbone MA, Zhu D, Anholt RRH, et al. (2012) Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci U S A* 109: 15553–15559.
32. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, et al. (2012) The *Drosophila melanogaster* genetic reference panel. *Nature* 482: 173–178.
33. Kacser H, Burns JA (1979) Molecular democracy: Who shares the controls? *Biochem Soc Trans* 7: 1149–1160.
34. Suarez RK, Moyes CD (2012) Metabolism in the age of ‘omes’. *J Exp Biol* 215: 2351–2357.
35. Fell DA (1992) Metabolic control analysis: experimental development of its theoretical and experimental development. *Biochem J* 286: 313–330.
36. Keightley PD (1989) Models of quantitative variation of flux in metabolic pathways. *Genetics* 121: 869–876.
37. 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: 565–569.
38. 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.
39. 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.
40. Zaitlen N, Kraft P (2012) What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. *Hum Genet* 131: 1655–1664.
41. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. *Am J Hum Genet* 90: 7–24.
42. The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661–678.
43. 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: 119–124.
44. Bloom JS, Ehrenreich IM, Loo WT, Lite TLV, Kruglyak L (2013) Finding the sources of missing heritability in a yeast cross. *Nature* 494: 234–237.
45. Kemper KE, Goddard ME (2012) Understanding and predicting complex traits: knowledge from cattle. *Hum Mol Genet* 21: R45–R51.
46. Haile-Mariam M, Nieuwhof G, Beard K, Konstantinov K, Hayes B (2013) Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle from genomic and pedigree data and implications for genomic evaluations. *J Anim Breed Genet* 130: 20–31.
47. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, et al. (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 45: 353–361.
48. Kendler KS (2013) What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. *Mol Psychiatry*. doi: 10.1038/mp.2013.50.
49. Visscher PM, Goddard ME (2010) Systems genetics: the added value of gene expression. *HFSP J* 4: 6–10.

Acknowledgments

We thank Eric Lander for critical comments on a previous version of the paper which stimulated what we believe are improvements in it. We acknowledge funding from the Australian Research Council (FT0991360) and National Health and Medical Research Council (1050218).

Author Contributions

Wrote the paper: SS EMD RSK WGH NRW.