A Two-State Epistasis Model Better Explains Complex Traits and Increases Power to Detect Additive Risk Variants

Kerry L. Bubb and Christine Queitsch

April 7, 2015

Department of Genome Sciences, University of Washington

Seattle, Washington 98115, USA

Running Head: Two-State Epistasis Model improves GWAS

Key Words: GWAS, Robustness, Epistasis, Heritability

Corresponding Authors:

Kerry L. Bubb

Christine Queitsch

Department of Genome Sciences

University of Washington

Box 355065

Seattle, WA 98195

(206) 685-8935 (ph.)

kbubb@u.washington.edu

queitsch@u.washington.edu

ABSTRACT

Despite more than a decade of effort, the genetic underpinnings of many complex traits and diseases remain largely elusive. It is increasingly recognized that a purely additive model, upon which most genome-wide association studies (GWAS) rely, is insufficient. Although thousands of significant trait-associated loci have been identified, they often explain little of the inferred genetic variance. Several factors have been invoked to explain the 'missing heritability', including epistasis. Accounting for all possible epistatic interactions is computationally complex and requires very large samples. Here, we propose a simple two-state epistasis model, in which individuals show either high or low variant penetrance with respect to a certain trait. The use of this model increases the power to detect additive trait-associated loci. We show that this model is consistent with current GWAS results and better fits heritability observations based on twin studies. We suggest that accounting for variant penetrance will significantly increase our power to identify underlying additive loci.

INTRODUCTION

We have been remarkably successful in identifying the genetic causes of Mendelian diseases with pedigree and sequence analysis (Botstein and Risch 2003; Gilissen et al. 2011). In contrast, the genetic underpinnings of multigenic heritable traits and diseases remain largely unknown. Many complex diseases such as autism, schizophrenia and cleft lip/palate have discrete, binary states (affected vs unaffected), yet these discrete states are determined by whether an individual passes some threshold in an underlying quantitative liability. This liability is both genetic and environmental in origin (Stefansson et al. 2014; Lynch and Walsh 1998).

GWAS is the most commonly used approach to discover the multiple loci underlying complex traits, relying on the assumption that risk alleles are more common among affected individuals than unaffected individuals. As currently implemented, GWA also assumes that loci contribute independently and additively to the total genetic contribution of a complex trait. How successful this experimental design has been is a matter of debate. Since 2007, GWAS have identified thousands of novel loci with significant trait or disease association and implicated previously unconnected pathways in disease processes (Visscher et al. 2012). Examples of clear biological and clinical relevance include the autophagy pathway in Crohn's disease (Rioux et al. 2007; Hampe et al. 2007; Prescott et al. 2007; Yamazaki et al. 2007) and the JAK-STAT signaling pathway in rheumatoid arthritis (Stahl et al. 2010; Diogo, Okada, and Plenge 2014), among others. However, many trait-associated loci explain little of the inferred genetic variance. They are therefore of limited use for

predicting the disease risk of specific individuals and for understanding the mechanistic underpinnings of complex traits. It is widely acknowledged that there is room for improvement (Wray, Goddard, and Visscher 2007; Wray et al. 2013).

Many hypotheses have emerged to explain the 'missing heritability' (Eichler et al. 2010; Manolio et al. 2009; Gibson 2011). Some complex diseases such as autism spectrum disorder represent likely insufficiently resolved pools of phenotypically similar, but inherently rarer, disease traits with different genetic underpinnings. If so, the observed odds ratios for significantly trait-associated SNP are low because the common, trait-associated SNP is linked to a rare causative mutation that only appears in a small subset of haplotypes (Mitchell 2012). Fine-grained phenotyping rather than relying on discrete, binary diagnoses should help to explore this hypothesis (Walter 2013). Some have suggested that the heritability of complex traits is overestimated (Wilson 2008; Zuk et al. 2012). Studies accounting for all SNPs genome-wide simultaneously, as opposed to individually associating SNPs with traits, indicate that this explanation is unlikely for many traits (Yang et al. 2010; Zaitlen et al. 2013). Others have invoked currently inaccessible genetic or structural variants or rare risk alleles of moderate effect as major factors in complex traits (McClellan and King 2010; Eichler et al. 2010; Manolio et al. 2009). However, at least for autism, recent studies suggest that common variants account for over 50% of the genetic risk (Klei et al. 2012; Gaugler et al. 2014).

Finally, although additive genetics is certainly a major fodder for evolution and selective breeding since even epistatic interactions present mainly as additive (James

F Crow 2010; Mäki-Tanila and Hill 2014), epistasis can have large influence on complex traits (Lachowiec et al. 2014). A highly socially relevant example for the importance of epistasis comes from plant breeding. As breeders increased seed yield, presumably via additive genetic factors, seeds became far more numerous, larger, and heavier. The increasing pressure on plant stalks required new mutations that enabled plants to remain erect under the increased seed weight -- this epistatic interaction enabled the Green Revolution (Hedden 2003) that vastly increased food security in many poor parts of the world (Ramanathan 2009).

In humans, twin concordance rates indicate that at least some of the genetic variation influencing complex diseases is non-additive. Experimental evidence demonstrates that epistasis, *i.e.* the phenotypically relevant, and often non-reciprocal interaction of non-allelic genes, is pervasive in complex traits in various model organisms (Lehner et al. 2006; Queitsch, Carlson, and Girirajan 2012; Mackay 2014). There is no reason to assume that the genetic architecture of complex traits differs between humans and other highly complex eukaryotes. Therefore, the inclusion of epistatic effects in statistical models has been increasingly suggested and even attempted in some studies (Lachowiec et al. 2014; Hemani et al. 2014a; Hemani et al. 2011). Although models that allow for all gene × gene (and gene × gene × gene, etc) interactions will be more realistic, such a higher-order models require much larger datasets and faster algorithms (Lippert et al. 2013; Hemani, Knott, and Haley 2013; Hemani et al. 2014a; Wood et al. 2014; Li, Horstman, and Chen 2011; Hemani et al. 2014b).

Studies in model organisms suggest that a simpler, two-state epistasis may apply to complex traits and diseases (Gibson 2009; Mackay 2014). Certain genes can act as strong genetic modifiers for many others loci resulting in highly significant 'gene x genome' interactions as opposed to the familiar, often small-effect, gene x gene interactions. Known examples include the chaperone Hsp90 in plants, flies, worms, fish, and yeast (Rutherford and Lindquist 1998; Jarosz and Lindquist 2010; Queitsch, Sangster, and Lindquist 2002; Casanueva, Burga, and Lehner 2012; Rohner et al. 2013) and several chromatin remodeling proteins in worms, among others (Lehner et al. 2006; Levy and Siegal 2008). In addition to increasing the penetrance of many genetic variants in genetically diverse populations, perturbation of strong genetic modifiers can also vastly increase phenotypic variation in isogenic populations (Richardson et al. 2013). Notably, the activity of strong genetic modifiers can be modulated by environmental stress (Rutherford 2003). Based on studies in plants, worms, and yeast, the number of strong genetic modifiers is small, possibly ~10% of all genes (Lehner et al. 2006; Levy and Siegal 2008; Fu et al. 2009).

There is some empirical support for strong genetic modifiers playing a role in complex diseases. For example, exome sequencing of autism parent-child trios finds disease-associated *de novo* mutations in CHD8, a chromatin remodeling factor involved in beta-catenin and p53 signaling (O'Roak, Vives, Girirajan, et al. 2012). Other disease-associated *de novo* mutations also implicate chromatin remodeling and beta-catenin signaling, which are both general developmental pathways, rather than specific neurodevelopmental pathways (O'Roak, Vives, Fu, et al. 2012; Krumm

et al. 2014). Chromatin remodeling is also implicated in schizophrenia, a distinct neurodevelopmental disease (Koga et al. 2009; McCarthy et al. 2014). It seems unlikely that perturbation of a general pathway like chromatin remodeling leads to two distinct diseases; rather, these perturbations may increase penetrance of disease-specific genetic variants. Similarly, network analysis of asthma GWAS hits and interactome data (genomics, gene expression, drug response) found three novel pathways that contribute to an 'asthma disease module': the AP1 family of transcription factors, Trka Receptor Signaling, and the GAB1 signalosome (Sharma et al. 2015). All three are general signaling pathways of growth and development, involving canonical MAP kinase components; their specific relevance to asthma is unclear. In short, complex diseases may arise through perturbed strong genetic modifiers which epistatically enhance the penetrance of many other variants genome-wide. Perturbations may occur either through mutation (as above) or environmental stress (Thomas 2010; Tsuang et al. 2004; Gibson 2009). Assuming a threshold model for complex diseases, this strong interaction would cause a large shift of variant effect across the disease threshold.

Although the existence of differences in variant penetrance among individuals is not a new concept, its effect on GWAS in humans has not been investigated. Here, we present a simple two-state epistasis model, in which disease status of an individual depends on a combination of additive alleles (as before) as well as their penetrance in a given individual. Such a model might be called an A×P, or Additive × Penetrance model. A population consisting of all individuals with increased

penetrance of many different genetic variants will have higher phenotypic variation, *i.e.* it will be less phenotypically robust, than an otherwise equivalent population consisting of all robust individuals with low penetrance (Queitsch, Carlson, and Girirajan 2012). Theoretical population genetics provides a strong argument for the potential benefits of maintaining a population with a balance of robust and non-robust individuals (Le Cunff and Pakdaman 2012; Hermisson and Wagner 2004; Huerta-Sanchez and Durrett 2007; Wagner 2008; Ferrada and Wagner 2008; Draghi et al. 2010; Wagner 2012). As we show, this model increases the power to detect additive trait-associated loci and better fits heritability observations based on twin studies. In the absence of robustness measures in humans, one may argue that such model is of limited use to improve GWAS in humans. However, we argue that our model's success should inform our approach to finding disease-causing loci and we discuss strategies how to apply it to existing and future GWAS data.

RESULTS

Heritability estimates imply substantial non-additivity of genetic factors contributing to human disorders: Geneticists refer to two types of heritability: additive (*narrow-sense* or h^2) and total (*broad-sense* or H^2). Since broad-sense heritability (H^2) is the fraction of phenotypic variance explained by *all* genetic factors, and narrow-sense heritability (h^2) is the fraction of phenotypic variance explained by *additive* genetic factors, it is tempting to deduce non-additive genetic effects (epistatic effects) by simply subtracting these two values. However, despite numerous

references to heritability in the literature, there remains some confusion on the methods of calculating H^2 in humans (*see Appendix I*) (Zuk et al. 2012; Hemani, Knott, and Haley 2013).

The additive heritability of *quantitative* traits is simply a function of the slope of the line of regression between pairs of related individuals (**Fig. 1A**) (Lynch and Walsh 1998; Visscher, Hill, and Wray 2008). However, complications arise for traits that are observably discrete, but rely on thresholding of an underlying, unobserved quantitative *liability* scale. Such traits are called *threshold characters* (Wright 1934; Rendel 1962; Lynch and Walsh 1998). Using the same method to measure additive heritability as before – doubling the slope of the line of regression between parent-offspring pairs – results in a substantially smaller estimate of the additive heritability (**Fig. 1A, inset**).

The degree to which this *observable* discrete additive heritability (h^2_o) is decreased relative to the heritability of the *underlying* additive liability (h^2) depends on where the diagnostic threshold is drawn – in other words, the fraction of individuals that are affected (φ). Henceforth, we use h^2_{bin} to refer to the observable heritability of threshold traits (binary, affected/unaffected).

In 1950, Dempster and Lerner (Dempster and Lerner 1950) derived a formula relating h^2 to h^2_{bin} using a simplifying assumption that the quantitative trait (liability) is normally distributed in the population (see *Appendix II*). Here, we refer to this Dempster-Lerner derivation as $T(h^2_{bin})$ as the maximum heritability of the binary

character attainable if the underlying liability consisted purely of additive genetic factors (h^2=1) (**Fig. 1B**).

Expected values for h^2_{bin} can also be determined via simulation, in which the additive liability is binomially distributed (*N* is twice the number of additive loci; *P* is the frequency of the risk allele at each locus). We refer to the maximum binary heritability attained via simulation of a *purely additive* model with $h^2=1$ as $S(h^2_{bin})$.

In either case, binary heritability h^2_{bin} reaches a maximum (with respect to h^2) when *half of the population is affected* and drops off rapidly as the fraction of affected individuals approaches values typical for common discrete complex traits, such as autism, schizophrenia, and multiple sclerosis (**Fig. 1B**). However, *empirically determined* h^2_{bin} values for these traits – $O(h^2_{bin})$ – are much higher (**Table 1**), calling a purely additive model into question.

To explore the implied non-additive factors, we developed a simulator that designates a subset of the population as non-robust by incorporating a robustness perturbation factor that increases the effect size of all additive risk alleles, *i.e.* increases variant penetrance. Using this simulator, a purely robust population maintains the theoretical relationship between the fraction of affected individuals (φ) and h^2_{bin} [here called $S(h^2_{bin})$] (**Fig. 1B**, black curve). The inclusion of non-robust individuals with increased variant penetrance results in an increase of h^2_{bin} [here called $S_r(h^2_{bin})$] even when the binary trait is at very low frequency (**Fig. 1B**, gold curve). Below, we explore the effect of varying this robustness factor, as well as the

frequency of non-robust individuals in the population and the contribution of non-genetic factors ($H^2<1$).

Hill and colleagues (Hill, Goddard, and Visscher 2008) proposed a simpler function to determine the extent of non-additive genetic effects: the correlation of monozygotic twins minus twice the correlation of dizygotic twins ($r_{MZ}-2r_{DZ}$). They use this statistic to evaluate both quantitative traits, such as height, and threshold traits, such as endometriosis. "$r_{MZ}>r_{DZ}$ implies that part of the resemblance is due to genetic factors and $r_{MZ}>2r_{DZ}$ implies the importance of non-additive genetic effects". This metric makes intuitive sense, because monozygotic twins share 100% their alleles, and dizygotic twins share 50% of their alleles. While most of the complex traits that Hill and colleagues examine in their paper had $r_{MZ}-2r_{DZ}$ values close to or less than zero, for many important complex diseases, $r_{MZ}-2r_{DZ} > 0$ (**Table 1**).

Therefore, rather than relying on traditional heritability metrics, we quantify the non-additive genetic component of complex traits and diseases by comparing *empirically determined* heritability, $O(h^2_{bin})$, to binary heritability with and without including a robustness perturbation factor [$S(h^2_{bin})$, additive model, assuming equal robustness in populations; $S_r(h^2_{bin})$, a two-state model, assuming differential robustness in population]. Simulated populations that are a mixture of robust and non-robust individuals produce levels of h^2_{bin} that are more consistent with empirically determined $O(h^2_{bin})$ from twin studies (**Table 1, Fig. 1C**).

Robustness as a two-state model of epistasis: Rather than expressing liability as a deviation from the mean, here, we define a quantitative phenotype *(y)* as a strictly positive value that is a linear combination of genetic factors and noise (environmental factors):

$$y = \sum_{i=1}^{n} \alpha_i g_i + \varepsilon \tag{1}$$

where the α_i's are the weights of each of the contributing genetic components, n is the number of contributing genetic components, ε is the residual effect, which is assumed to be normally distributed with mean zero and standard deviation σ_ε, and g_i is an indicator variable, taking values 0 or 1 depending on whether the risk allele is present at the *i*th locus. (**Fig. 2A**) The bi-allelic states can be thought of as intact (wild type) or broken (risk) versions of the gene. With exome sequencing, such data are now readily available (O'Roak et al. 2011). For this and all further models (including populations consisting of a mixture of subpopulations), we determine binary trait status of each individual using a threshold drawn on the empirical population of y, such that the fraction of affected individuals in the population is consistent with that of a given trait of interest (φ).

Note that if we assume that all risk alleles contribute equally to disease risk, (all α_i's are equal), the formula simplifies further to:

$$y = n\alpha + \varepsilon \tag{2}$$

where n is the number of risk-associated alleles.

A non-robust subpopulation could be defined in several ways. We explored four common manifestations of large-effect factors, represented by coefficients c_L, c_ε, c_α, and c_n that modify the base additive formula (**Fig. 2**), and evaluated their potential for explaining observed information. Note that at this point, we make no assumptions about whether robustness status is genetic or environmental in origin.

The first way in which the simple additive model could be modified is by inclusion of a large factor that simply adds to the phenotypic liability independently of other additive factors (**Fig. 2B**), where $c_L >> \alpha$. In this case the factor (c_L) is not epistatic to the other additive factors (g_i), it is yet another, larger, additive factor. These types of factors are relatively simple to detect with linkage analysis (if genetic) or epidemiological studies (if environmental). Examples include a mutation in the fibroblast growth factor receptor 3, which causes a decrease in height (achondroplasia) (Shiang et al. 1994; Rousseau et al. 1994), and William's syndrome, which causes, among other things, a decrease in IQ (Mervis and Becerra 2007). In both these cases, the standard deviation of the trait (height and IQ, respectively) stays the same as that for typically developing populations. A population with a novel factor of large effect, as described above, *would* have a different phenotypic coefficient of variation, a common metric for robustness (Lempe et al. 2005), than a population without that factor. However the difference is driven exclusively by a change in *mean* phenotype, and the novel factor *would not* be considered as having a robustness effect. We therefore exclude this model from further consideration.

A second modification to the simple additive model, and one that *would* result in an increase in phenotypic variance, is one in which the degree of stochastic noise (ε) in determining the phenotype is increased, (**Fig. 2C**, where for robust individuals, $c_\varepsilon=1$ and for non-robust individuals, $c_\varepsilon>1$). Such a change would affect the variance of the phenotypic value in the population without changing the mean value, and therefore such a large-effect factor would be considered a *robustness* factor. However, because c_ε does not depend on the additive factors (g_i), there is no *epistatic* effect. Note that this is the same as decreasing the importance of genetics in determining the phenotype (*i.e.* decreasing overall heritability). An example of a single gene perturbation affecting trait robustness is the *elf4* mutation that greatly increases variance of the plant circadian period but maintains mean value (Doyle et al. 2002). Another example would be the increased variation in normalized eye and orbit size caused by perturbation of Hsp90 in surface-dwelling *Astyanax mexicanus* fish, which did not change trait means (Rohner et al. 2013). As this type of epistasis decreases the importance of genetics, is unlikely to underlie complex diseases with high observed heritability and monozygotic twin concordance, and we therefore exclude it from further consideration.

The final two types of single-factor epistatic effects are (i) those that increase the effect size (α) of already-present additive alleles (**Fig. 2D**; c_α), and (ii) those that reveal *cryptic* risk-alleles, increasing the total number of loci (n) influencing the trait (**Fig. 2E**, c_n). In each scenario, populations consisting exclusively of either robustness type (robust individuals: $c_\alpha=1$ and $c_n=1$; non-robust individuals: $c_\alpha>1$ or $c_n>1$) would

differ in both the mean and the variance of the phenotype, and thus would satisfy the classical definition of robustness (Queitsch, Carlson, and Girirajan 2012). Because the robustness factor interacts with the additive factors (g_i), c_α and c_n would also be considered epistatic effects. We call these coefficients c_α and c_n *robustness perturbation factors*.

Although the simplified versions of the equations for the two scenarios (**Fig. 2D, E**), wherein all weights are equal, are equivalent: $y=(c_\alpha\alpha)n+\varepsilon$ and $y=\alpha(c_\alpha n)+\varepsilon$, we can distinguish them experimentally. First, trait mapping in non-robust individuals will reveal additional loci, compared to a comparable study in robust individuals, if loss of robustness exposes cryptic loci (Sangster, Salathia, Lee, et al. 2008; Sangster, Salathia, Undurraga, et al. 2008; Jarosz and Lindquist 2010; Rutherford and Lindquist 1998). Second, loss of robustness can increase effect size (*i.e.* penetrance) of individual mutations (Casanueva, Burga, and Lehner 2012; Burga, Casanueva, and Lehner 2011). Third, methods based on comparing parent phenotypic variation to sib-sib phenotypic variation have been developed to estimate the number of additive loci affecting a trait (Slatkin 2013; Penrose 1969). Under a cryptic genetic variation scenario, an all-robust population would appear to have more additive loci affecting a trait than an all-non-robust population, whereas under an increased effect size scenario, they would be the same.

Therefore, we implement these last two models, D&E, in our simulator (**Fig. 3**). Our aim was to measure the effect of changing certain parameter values, such as n, φ, c_α *and* c_n on expected observable values, such as λ_{MZ}, λ_{DZ}, and the odds ratio (*OR*)

of a locus found in a typical GWAS, while holding noise (ε) constant. Using a set of input parameters, we first simulate a set of families each containing two parents and a primary child (trio). We then simulate monozygotic and dizygotic twins of that primary child, and assign quantitative phenotypes (y) to each child. To do so, we sample the number of risk alleles per child, given the number of relevant loci and frequency of risk alleles at each locus. Third, children are designated to be robust or non-robust, based on the frequency of perturbed 'robustness' alleles, p_r. If a child is non-robust, its additive risk is multiplied by c_α or c_n and the noise factor ε is added (y). Fourth, using the liability (y) of primary children, we designate the top φ fraction (*e.g.* 1%) as affected and thereby define a risk threshold for affected and unaffected individuals. This threshold is used to determine affected/unaffected status of twins. Fifth, we calculate the monozygotic and dizygotic concordance rates (λ_{MZ} and λ_{DZ}, respectively), and perform *in silico* GWAS on affected vs unaffected primary children (not including simulated twins) to determine the odds ratio of individual additive risk loci (OR_{add}).

Our first result confirms intuition: in a mixed population, consisting of both robust ($c_\alpha=1;\ c_n=1$) and non-robust ($c_\alpha>1;\ c_n>1$) individuals under either model D or E, individuals crossing the clinical threshold will be disproportionately from the non-robust subgroup. As c_α or c_n increase, total liability (y) increases. This is true for both the non-robust subpopulation *and* the entire population, which is why we use a percentile (φ) rather than an absolute threshold to determine affected/unaffected status.

Secondly, we find that a population in which just 1% of individuals are non-robust can easily produce the range of empirically determined heritabilities of binary threshold traits [**Fig. 1B inset**, $S_r(h^2_{bin})$] and reported twin concordances and broad-sense heritabilities of several complex diseases (**Fig. 1C, Table 1**). This is due to the fact that as c_α or c_n increase and ε remains constant, the fraction of liability that is determined by genetics (H^2) increases for the non-robust subpopulation and, by extension, for the entire population.

Controlling for robustness status adds power to GWAS: Regardless whether robustness is modeled as hiding cryptic variation or reducing penetrance of variants, if the goal is to find additive risk-loci (*i.e*. which may be good therapeutic targets), it is best to use *only* robust individuals for both affected and unaffected groups (**Fig. 4B**). Of course, controlling for a hidden robustness state, which modifies the effect of many alleles, returns the experiment to the situation for which GWAS was designed -- a purely additive model. It is less intuitive why using only robust (rather than all non-robust) individuals improves our ability to detect additive risk alleles in GWAS, which is a result we explore below.

In the case of cryptic variation, robust individuals have fewer available additive loci than non-robust individuals, so there are fewer ways to cross the threshold number of risk-alleles; the risk alleles are concentrated at the non-cryptic loci. Alternatively, affected robust individuals may carry additive risk alleles of larger effect

size, *i.e.* a different type of risk alleles than non-robust affected individuals. Here we assume that non-robust and robust individuals carry the same additive alleles.

In the case of increased penetrance, the explanation is similar to that given for why common risk alleles are more easily found than rare alleles *that increase risk by the same amount* (Gibson 2011): effect size is a function not only of the case:control ratio of allele frequency, but also of the *magnitude* of the frequency of the alleles in affected and unaffected individuals (**Fig. 4A**). If all affected and unaffected individuals are robust (as compared to non-robust), more risk alleles are required to cross the threshold (for affected individuals) and more risk alleles are allowed for individuals *not* crossing the threshold (for unaffected individuals). Therefore, while the ratio of allele frequencies between affected and unaffected individuals is the same for both all robust and all non-robust groups, the magnitude of the allele frequencies is different, making those loci easier to find in robust populations.

Robustness status may be genetically-determined, but difficult to pinpoint: We state earlier that robustness status need not be genetically determined; however, we wish to explore reasonable scenarios in which it is.

Were robustness status encoded by a single genetic factor, this factor would be readily discernible, either by GWAS or linkage analysis because, under our model, affected individuals in mixed populations are disproportionately non-robust. However, robustness status is highly unlikely to be encoded by a single gene. Model organism studies suggest that a significant fraction of total number of genes – possibly up to

10% – can affect robustness (Lehner et al. 2006; Levy and Siegal 2008; Fu et al. 2009).

To model the scenario of multiple 'robustness' genes, we use a simple yet plausible house-of-cards model of robustness, in which all n_r 'robustness' loci must be functional for an individual to be robust. We observe that the odds ratio for any one 'robustness' locus (OR_r) decreases to within the range often seen in GWAS as $n_r>50$ (**Fig. 4C**). In other words, since there many possible 'robustness' loci, specific ones will be hard to find in GWAS.

With this result in mind, it may appear that models D and E are similar to the model that includes an additional additive factor of large effect (model B) in that affected individuals are disproportionately either non-robust (models D&E) or carriers of a large-effect alleles (model B) – particularly since multiple factors of large additive effect that result in indistinguishable phenotypes could conceivably work in a house-of-cards model. However, we argue that models D and E explain the marginal penetrance of some seemingly large effect factors and concomitant lack of Mendelian inheritance in pedigrees (Rosenfeld et al. 2013) as well as the presence of the multitude of other lesser risk-associated loci much more readily than model B.

DISCUSSION

We conclude that (i) a model that includes a house-of-cards robustness state explains the observed data for several complex diseases better than one without, (ii) when looking for loci with additive risk alleles it is best to use all robust individuals for

both cases and controls, and (iii) ‘robustness’ loci are unlikely to be identified in GWAS.

In any given family, a faulty ‘robustness’ allele passed on from parent to child will act as a large-effect risk allele with Mendelian inheritance. This explains both high concordance among relatives and missing heritability in GWAS.

An obvious challenge is how to identify robust and non-robust humans; this challenge is met in several ways in model organisms (Queitsch, Carlson, and Girirajan 2012; Lempe et al. 2005). In humans, one may identify ‘robustness’ genes by comparing individuals with high comorbidity of complex diseases to those with none; however, to our knowledge, this has yet to be done. Another currently available approach would be to pool individuals that are affected by distinct complex diseases and compare these to their pooled unaffected controls. As we expect non-robust individuals to be overrepresented among affected individuals for any complex disease, this GWAS approach may identify ‘robustness’ loci because the frequency of perturbed ‘robustness’ loci is increased. Indeed, this approach has shown promise for neurological disorders (“Identification of Risk Loci with Shared Effects on Five Major Psychiatric Disorders: A Genome-Wide Analysis” 2013). Similarly, the distinct diseases autism spectrum disorder and schizophrenia converge on chromatin remodeling as a general pathway that is disrupted in affected individuals (O’Roak, Vives, Girirajan, et al. 2012; O’Roak, Vives, Fu, et al. 2012; Krumm et al. 2014; Koga et al. 2009; McCarthy et al. 2014). We suggest that perturbation of chromatin remodeling leads to increased penetrance of different additive risk alleles (*i.e.* at different loci), hence resulting in

distinct diseases. In fact, perturbation of chromatin remodeling genes increases the penetrance of many genetic variants in worms (Lehner et al. 2006).

The cited autism studies also point to a promising path towards identifying 'robustness' loci: whole-genome sequencing of parent-child trios (unaffected parents, affected child) (O'Roak et al. 2011; O'Roak, Vives, Girirajan, et al. 2012). This approach will capture the hypothesized large-effect risk alleles, which are either inherited in families or arise *de novo* in affected children. Notably, exome sequencing of autism parent-child trios finds significantly associated *de novo* mutations in CHD8, a chromatin remodeling factor with roles in beta-catenin and p53 signaling (O'Roak, Vives, Fu, et al. 2012). Other *de novo* mutations in affected children also implicate chromatin remodeling and beta-catenin signaling, both general developmental pathways, rather than specific neurodevelopmental pathways (O'Roak, Vives, Fu, et al. 2012; Krumm et al. 2014).

A third approach possible with current data would be to assume that the GWAS loci associated with the highest disease risk represent 'robustness' genes – particularly if there is no obvious association between gene function and a specific disease. One would then proceed by controlling for these additive risk allele. As a community, we have been reluctant to stray from additive models because of the rapidly rising complexity involved in accounting for interactions between genes. Also, multiplicative models assuming many interactions will yield lower concordance rate between first-degree relatives than is typically observed. This approach contrasts with ours as we assume 'gene x genome' epistasis (*i.e.* strong genetic modifiers

epistatically interact with many other loci). Note that statistical additivity is often an emergent property of underlying epistatic interactions (Song, Wang, and Slatkin 2010; Mackay 2014; Mäki-Tanila and Hill 2014). Nevertheless, from the viewpoint of developing predictive diagnostics and effective therapeutics, it is important to identify the mechanistic underpinnings of complex diseases. If 'gene x genome' epistasis contributes to complex disease as we suggest, we need to think of ways to identify 'robustness' genes or markers for robustness in humans.

One plausible proxy for robustness could be the level of genome-wide heterozygosity. For many traits in many organisms F_1 hybrids show less phenotypic variance and indeed hybrid vigor compared to their parental inbred lines P_A and P_B (Robertson and Reeve 1952; Festing 1976; Becker and Léon 1988; Phelan and Austad 1994; Klempt et al. 2006). Plant breeders have been using this simple principle for almost a century (J F Crow 1998). Hybrids show decreased phenotypic variance and increased vigor despite the fact that all three populations (F_1, P_A and P_B) are isogenic and therefore all phenotypic variation *within* each population should be environmental in origin. Because phenotypic variation due to environment is apparently reduced by heterozygosity, the *fraction* of phenotypic variation due to genetics should be greater in increasingly heterozygous populations. Hence, a good proxy for robustness status may be a genome-wide heterozygosity score (Campbell et al. 2007; Govindaraju et al. 2009; Lie, Simmons, and Rhodes 2009; Gage et al. 2006), which is often readily calculated from the same type of data collected for GWAS.

The relationship between heterozygosity and disease is reflected in the sex bias of neurodevelopmental disorders (Jacquemont et al. 2014), with many more males than females affected (~30% to 50% excess, autistic male/female ratio of 4:1). Autistic females tend to carry a higher load of copy number variants (CNV) and single nucleotide variants (SNV) than affected males, arguing that a greater genetic liability is required for females to cross the disease threshold. A likely explanation for this “female protection” is the second X chromosome, which ‘buffers’ otherwise deleterious mutations (Jacquemont et al. 2014). Moreover, autistic females are often much more severely affected; many fewer girls (one in seven) than boys have highly functioning ASD (autism spectrum disorder). Taken together, increased ASD severity, decreased occurrence, and higher mutational burden in females strongly implicate non-additive genetic effects in ASD.

Ideally, we would have readily accessible and affordable DNA-, RNA-, protein- or cell-based markers to determine robustness levels across many individuals and apply robustness as a covariate in GWAS. Without extensive studies in model organisms, which offer orthogonal measures to determine robustness (Lempe et al. 2005), the identification of such markers seems not easily feasible. Some have suggested that levels of somatic genetic variation may be a read-out of robustness and mutation penetrance with higher levels of somatic variation indicating lower robustness (Queitsch, Carlson, and Girirajan 2012; Press, Carlson, and Queitsch 2014; Heng 2010). Single-cell sequencing and phenotyping may offer insights into an individual’s robustness with greater cell-to-cell variation indicating lower robustness and higher

mutation penetrance (Steininger et al. 2014; Altschuler and Wu 2010). Simpler alternatives may rely on identifying individuals that are outliers in expression or methylation. Similar to using genotype data to elucidate population structure (Pritchard, Stephens, and Donnelly 2000), expression (and methylation) data could be used to find subpopulations with different patterns of covariance between gene expression (methylation) levels (Satoh et al. 2006). Admittedly, none of these suggestions are exactly affordable; yet, compared to the vast resources committed to GWAS, exploring these potential markers in model organism studies seems worthwhile. Notably, as robustness status can be environmentally-determined, we believe that such markers would detect environmentally-dependent robustness defects. Although a single environmental risk-factor should be readily identified through epidemiological studies, there could be a time lag between the environmental insult and the disease or a house-of-cards mechanism for multiple environmental insults that make them difficult to pinpoint. Even without knowing the etiology of robustness status, we need to separate individuals into robust and non-robust categories.

Are there indeed robust and non-robust individuals as we argue? Population genetics theory suggests that it is evolutionarily advantageous to maintain a balance of robust and non-robust individuals within a population, mainly due to the fact that non-robust individuals supply a broader phenotypic range on which selection can act under extreme circumstances (Le Cunff and Pakdaman 2012; Hermisson and Wagner 2004; Huerta-Sanchez and Durrett 2007; Wagner 2008; Ferrada and Wagner 2008; Draghi et al. 2010; Wagner 2012).

Medical professionals intuitively agree. Drawing on their experience, they evaluate the full "Gestalt" of a patient and predict a patient's risk for a negative outcome with often great precision. Applying 'robustness' markers may turn a physician's intuition into diagnostics. Because of the potential for increasing predictive power and identifying effective drug targets, the possibility that robustness is a major player in the etiology of complex disease should be carefully considered.

APPENDIX

(I) Broad-sense, or Total Heritability: The definition of total, or *broad-sense* heritability is straightforward:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\epsilon^2} \tag{3}$$

where σ_g^2 is the phenotypic variance due to genetic factors and σ_ϵ^2 is the phenotypic variance due to environmental factors. In model organisms, such as yeast, these values can be measured directly using replicates (Bloom et al. 2013). In humans, however, we must rely on an often slim set of naturally-occurring replicates – monozygotic twins.

Multiple different methods of calculating total heritability are still used, most of which are simple functions of monozygotic and dizygotic similarity, but the most common was proposed by Holzinger in 1929 (Holzinger 1929):

$$H^2 = \frac{r_{MZ} - r_{DZ}}{1 - r_{DZ}} \quad (4)$$

where r_{MZ} and r_{DZ} stand for the correlation between the phenotypes of pairs of monozygotic and dizygotic twins, respectively.

Adding to the complexity of these calculations is the fact that traits can be either continuous (quantitative) or discrete (affected or unaffected). In the case of Holzinger's formula for total heritability (H^2), the same formula can be used. The following is a useful simplification of the original formula for the case of discrete traits, where unaffected and affected individuals assume values of 0 and 1 respectively:

$$H^2 = \frac{(1-\lambda_{DZ}) - (1-\lambda_{MZ})}{1-\lambda_{DZ}} \quad (5)$$

where λ_{MZ} and λ_{DZ} are concordance rates between monozygotic and dizygotic twins, respectively.

Despite known problems surrounding the calculation of broad-sense heritability, specifically over-simplifying assumptions of (i) non-linkage between risk loci, and (ii) similarity of environment for monozygotic and dizygotic twins (Jacquard 1983; Schönemann 1997; Lynch and Walsh 1998), the above formula for H^2 remains a common method for calculating the "heritability" reported for many complex diseases. To side-step confusion surrounding heritability calculations, we use simple monozygotic and dizygotic twin concordance as an output measurement.

Other methods of calculating include Falconer's formula, $2(r_{MZ} - r_{DZ})$ and Nichols's formula $2(r_{MZ} - r_{DZ})/r_{MZ}$, however these methods clearly fail in an essential

criteria of the H^2 statistic: that H^2 *must* have a value between zero and one, as it is, after all, a fraction – specifically, the fraction of phenotypic variance due to genetic effects. For example, for Autism Spectrum Disorders, the concordance of monozygotic twins is 0.95 and the concordance of dizygotic twins is 0.04, and an overall population prevalence of 0.5 implies r_{MZ}=0.97 and r_{DZ}=0.20, which would make H^2 1.55 and 1.59 for Falconer's and Nichols's formulas, respectively. It is safe to say that calculating the broad-sense heritability in humans is an intractable problem.

(II) Narrow-sense, or Additive Heritability of Threshold Traits: The relationship between the observable heritability of a discrete threshold trait (h^2_{bin}) and the additive heritability of the underlying, often hidden quantitative trait (h^2) was derived as follows by Dempster and Lerner in 1950 (Dempster and Lerner 1950; Lynch and Walsh 1998), under the assumption that the underlying additive liability is normally distributed:

$$h^2_{bin} = h^2 \{ \frac{[p(x_p)]^2}{\phi_p(1-\phi_p)} \} \tag{6}$$

where $p(x_p)$ is the height of the standardized normal distribution at the diagnostic threshold (above which, an individual would be "affected") and φ_p is the fraction of affected individuals in the population. h^2_{bin} reaches a maximum value of $2/\pi \approx 0.637$ when *all* of the underlying liability is genetic and additive ($h^2=1$) and half of the population is affected ($\varphi=0.5$). When $\varphi=0.01$, a typical value for more common discrete complex traits such as autism and schizophrenia, observable heritability

(h^2_{bin}) based purely on underlying *additive* genetic liability is expected to be 0.072. The maximum value for h^2_{bin} decreases as the fraction of the population affected φ decreases.

In our simulations, genetic liability is described by a binomial distribution, with the number of trials being twice the number of risk loci, or *2n* (for a diploid individual) and the probability of success being the frequency of a risk allele at any given locus, or *p*. Using this distribution, we get values of h^2_{bin} that are slightly higher than those predicted by Dempster and Lerner based on a normal distribution for all values of φ other than a small window around $\varphi=0.5$ (roughly $0.4 < \varphi < 0.6$) (**Fig. 1B**, **Table 1**).

ACKNOWLEDGMENTS

We thank Joe Felsenstein for providing insightful advice that shaped this the manuscript. We thank Mary Kuhner, Cristina Alexandre. Joshua Schraiber, and Yaniv Erlich for helpful comments and discussions, and Alex Mason for help preparing the manuscript. This work was supported by a grant the National Institute of Health (DP2OD008371).

Figure 1: Comparison of *narrow-sense* heritability of quantitative liability, *observable* heritability of discrete traits, and *empirically determined* heritability of certain complex threshold traits. (**A**) Each symbol in the large scatter plot represents the genetic liabilities for one parent-child pair. Each symbol in the inset scatter plot represents the disease state for one parent-child pair (disease state must be 0=unaffected or 1=affected; points were jittered for visual clarity). Green pluses indicate cases in which the child is affected (*i.e.* has crossed the diagnostic threshold) and the parent is not affected; blue crosses indicate cases in which the parent is affected and the child is not; red filled circles indicate cases in which both are affected; black open circles indicate cases in which neither is affected. Input parameter values for the simulation were as follows: number of additive risk-loci (n) = 100; frequency of additive risk-allele (p) = 0.1; frequency of non-robust state (p_r) = 0. (**B**) The relationship between frequency of affected individuals in the population (φ) and the maximum possible observable *binary* heritability (h^2_{bin}) when the underlying liability is entirely determined by additive genetic factors (h^2 = 1) as predicted by theory (Dempster and Lerner 1950), $T(h^2_{bin})$, as determined by simulation, $S(h^2_{bin})$, or as determined by simulation of a population that includes two types of individuals, robust and non-robust, at a ratio of 99:1, $S_r(h^2_{bin})$, where H^2 = 1. In the inset, we display the relevant φ range where 1% or less of the population is affected, which is common for many complex diseases. For the diseases shown, empirically determined heritability $O(h^2_{bin})$ is up to twice as high as predicted under the models that do not include an epistatic robustness factor. (**C**) In a population containing just 1% non-robust individuals, increasing the robustness perturbation factor (c_α), *i.e.* the fold-change of the additive genetic liability, increases both heritability and twin concordance to levels observed in several complex diseases.

Figure 2: Base additive model and modifications. (**A**) Base additive model. (**B**) Additive model plus single large-effect factor (c_L). (**C**) Additive model with increased noise, *i.e.* increased residual effect (c_ε). (**D**) Additive model assuming existence of non-robust state that increases effect size of additive risk alleles (c_α). (**E**) Additive model assuming existence of non-robust state that reveals previously phenotypically silent (cryptic) risk alleles (c_n), represented by the additional blue triangles. [Note: y is some quantitative phenotype, α_i's are the weights of each of the contributing genetic components, n is the number of contributing genetic components, c_ε is the residual effect, which is assumed to be normally distributed with mean zero and standard deviation σ_ε, and g_i is an indicator variable, taking values 0 or 1 depending on whether the risk allele is present at the i-th locus.]

Figure 3: Simulator schematic. Using given *INPUT* parameters, we performed a simulation to generate a set of families consisting of two parents, a primary child (in bold box), and both a monozygotic and a dizygotic twin for each primary child. Each individual was assigned two alleles at random (either a risk allele (red), with probability p, or a non-risk allele, with probability $1-p$) for each of the n loci simulated. Each individual was also assigned a robustness status in the following way: if either of the two alleles at any of the n_r 'robustness' loci were nonfunctional, each with probability p_r, the individual was non-robust (indicated by gold diamond). Trait threshold is determined such that a fraction φ of primary children is affected. We then performed twin concordance calculations (using the simulated twins) and GWAS (using only primary individuals) to produce a set of *OUTPUT* values that we compared to empirically observed values.

Figure 4: Effect size of an additive risk locus is determined by both the total number of additive risk loci and the presence/absence of 'gene x genome' epistasis. (**A**) In a robust population, where genetic liability is fully determined by the number of additive alleles, both affected and unaffected individuals will have more risk alleles than individuals of a non-robust population. Therefore the odds ratio (OR_{add}) for any given additive allele in the robust population (2.33) will be greater than in non-robust populations (2.06), even if risk allele ratios are the same (2). (**B**) Within a population consisting of 99% robust (model A) and 1% non-robust (model D or E), the effect size of an additive allele is highest when GWAS is performed in a subpopulation consisting of only robust individuals (black line) and lowest when GWAS is performed in a subpopulation sampled without regard to robustness state (green line), with subpopulations consisting of only non-robust individuals of either type D or E performing at intermediate levels (blue lines). Input parameter values for the simulations were as follows: number of additive risk loci $20 \leq (n) \leq 200$; frequency of additive risk allele (p) = 0.02; frequency of non-robust state (p_r) = 0.01; robustness perturbation factor (c_α) = 2. (**C**) Increasing the number of genes in which perturbation causes a non-robust state decreases the effect size of any one such 'robustness' gene to levels comparable to that most commonly found for additive risk alleles in which the risk allele is the minor allele ($1 < OR_{buff} < 2$, indicated by grey horizontal bar). Input parameter values for the simulations were as follows: number of additive risk-loci (n) = 100; frequency of additive risk-allele (p) = 0.02; frequency of non-robust state (p_r) = 0.01; robustness perturbation factor $1 \leq c_\alpha \leq 3$.

Table 1: Heritability of select Complex Traits

Trait	λ_{MZ}	λ_{DZ}	φ	$T(h^2_{bin})$	$S(h^2_{bin})$	$O(h^2_{bin})$	r_{MZ} - $2r_{DZ}$	H^2
Autism Spectrum Disorder	0.92 (Bailey et al. 1995)	0.1 (Bailey et al. 1995)	0.5%	0.042	0.1	0.2	0.33	0.91
Schizophrenia	0.4-0.7 (Onstad et al. 1991; Cardno and Gottesman 2000)	0-0.26 (Onstad et al. 1991; Cardno and Gottesman 2000)	1%	0.072	0.16	0.26	0.11	0.44
Multiple sclerosis	0.24 (Hansen et al. 2005)	0.03 (Hansen et al. 2005)	0.1%	0.011	0.04	0.06	0.14	0.22
Type I diabetes	0.53 (Kyvik, Green, and Beck-Nielsen 1995)	0.11 (Kyvik, Green, and Beck-Nielsen 1995)	0.3%	0.028	0.07	0.22	0.07	0.47
Cleft Lip/Palate	0.50 (Grosen et al. 2011)	0.08 (Grosen et al. 2011)	0.15%	0.016	0.06	0.16	0.14	0.46

λ_{MZ} = monozygotic twin concordance; λ_{DZ} = dizygotic twin concordance; φ = frequency in population; $T(h^2_{bin})$ = maximum h^2_{bin} attainable when $h^2 = 1$ and assuming that the liability of the trait follows a normal distribution (Dempster and Lerner 1950); $S(h^2_{bin})$ = h^2_{bin} attained when $h^2 = 1$ and simulating the liability of the trait influenced by 100 additive risk loci as a binomial distribution, where $N = 2\times100$ and $p = 0.5$; $O(h^2_{bin})$ = observed heritability of a binary threshold trait; $r_{MZ} - r_{DZ}$ is the statistic used by Hill and colleagues (Hill, Goddard, and Visscher 2008) [*Note: r* (*correlation*) is related but not identical to λ (*concordance*).]; H^2 is Holzinger's broad-sense heritability (Holzinger 1929).

LITERATURE CITED

Altschuler, Steven J, and Lani F Wu. 2010. “Cellular Heterogeneity: Do Differences Make a Difference?” *Cell* 141 (4): 559–63.

Bailey, A, A Le Couteur, I Gottesman, P Bolton, E Simonoff, E Yuzda, and M Rutter. 1995. “Autism as a Strongly Genetic Disorder: Evidence from a British Twin Study.” *Psychological Medicine* 25 (1): 63–77.

Becker, H C, and J Léon. 1988. “Stability Analysis in Plant Breeding.” *Plant Breeding = Zeitschrift Fur Pflanzenzuchtung* 101 (1): 1–23.

Bloom, Joshua S, Ian M Ehrenreich, Wesley T Loo, Thúy-Lan Võ Lite, and Leonid Kruglyak. 2013. “Finding the Sources of Missing Heritability in a Yeast Cross.” *Nature*, February. doi:10.1038/nature11867.

Botstein, David, and Neil Risch. 2003. “Discovering Genotypes Underlying Human Phenotypes: Past Successes for Mendelian Disease, Future Approaches for Complex Disease.” *Nature Genetics* 33 Suppl (March): 228–37.

Burga, Alejandro, M Olivia Casanueva, and Ben Lehner. 2011. “Predicting Mutation Outcome from Early Stochastic Variation in Genetic Interaction Partners.” *Nature* 480 (7376): 250–53.

Campbell, Harry, Andrew D Carothers, Igor Rudan, Caroline Hayward, Zrinka Biloglav, Lovorka Barac, Marijana Pericic, et al. 2007. “Effects of Genome-Wide Heterozygosity on a Range of Biomedically Relevant Human Quantitative Traits.” *Human Molecular Genetics* 16 (2): 233–41.

Cardno, Alastair G, and Irving I Gottesman. 2000. “Twin Studies of Schizophrenia: From

Bow-and-Arrow Concordances to Star Wars Mx and Functional Genomics." *American Journal of Medical Genetics* 97 (1): 12–17.

Casanueva, M Olivia, Alejandro Burga, and Ben Lehner. 2012. "Fitness Trade-Offs and Environmentally Induced Mutation Buffering in Isogenic C. Elegans." *Science* 335 (6064): 82–85.

Crow, J F. 1998. "90 Years Ago: The Beginning of Hybrid Maize." *Genetics* 148 (3): 923–28.

Crow, James F. 2010. "On Epistasis: Why It Is Unimportant in Polygenic Directional Selection." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 365 (1544): 1241–44.

Dempster, E R, and I M Lerner. 1950. "Heritability of Threshold Characters." *Genetics* 35 (2): 212–36.

Diogo, Dorothée, Yukinori Okada, and Robert M Plenge. 2014. "Genome-Wide Association Studies to Advance Our Understanding of Critical Cell Types and Pathways in Rheumatoid Arthritis: Recent Findings and Challenges." *Current Opinion in Rheumatology* 26 (1): 85–92.

Doyle, Mark R, Seth J Davis, Ruth M Bastow, Harriet G McWatters, László Kozma-Bognár, Ferenc Nagy, Andrew J Millar, and Richard M Amasino. 2002. "The ELF4 Gene Controls Circadian Rhythms and Flowering Time in Arabidopsis Thaliana." *Nature* 419 (6902): 74–77.

Draghi, Jeremy A, Todd L Parsons, Günter P Wagner, and Joshua B Plotkin. 2010. "Mutational Robustness Can Facilitate Adaptation." *Nature* 463 (7279): 353–55.

Eichler, Evan E, Jonathan Flint, Greg Gibson, Augustine Kong, Suzanne M Leal, Jason H Moore, and Joseph H Nadeau. 2010. "Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease." *Nature Reviews. Genetics* 11 (6): 446–50.

Ferrada, Evandro, and Andreas Wagner. 2008. “Protein Robustness Promotes Evolutionary Innovations on Large Evolutionary Time-Scales.” *Proceedings. Biological Sciences / The Royal Society* 275 (1643): 1595–1602.

Festing, Michael F W. 1976. “Phenotypic Variability of Inbred and Outbred Mice.” *Nature* 263 (5574): 230–32.

Fu, Jingyuan, Joost J B Keurentjes, Harro Bouwmeester, Twan America, Francel W A Verstappen, Jane L Ward, Michael H Beale, et al. 2009. “System-Wide Molecular Evidence for Phenotypic Buffering in Arabidopsis.” *Nature Genetics* 41 (2): 166–67.

Gage, Matthew J G, Alison K Surridge, Joseph L Tomkins, Emma Green, Louise Wiskin, Diana J Bell, and Godfrey M Hewitt. 2006. “Reduced Heterozygosity Depresses Sperm Quality in Wild Rabbits, Oryctolagus Cuniculus.” *Current Biology: CB* 16 (6): 612–17.

Gaugler, Trent, Lambertus Klei, Stephan J Sanders, Corneliu A Bodea, Arthur P Goldberg, Ann B Lee, Milind Mahajan, et al. 2014. “Most Genetic Risk for Autism Resides with Common Variation.” *Nature Genetics* advance online publication (July). doi:10.1038/ng.3039.

Gibson, Greg. 2009. “Decanalization and the Origin of Complex Disease.” *Nature Reviews. Genetics* 10 (2): 134–40.

———. 2011. “Rare and Common Variants: Twenty Arguments.” *Nature Reviews. Genetics* 13 (2): 135–45.

Gilissen, Christian, Alexander Hoischen, Han G Brunner, and Joris A Veltman. 2011. “Unlocking Mendelian Disease Using Exome Sequencing.” *Genome Biology* 12 (9): 228.

Govindaraju, Didahally R, Martin G Larson, Xiaoyan Yin, Emelia J Benjamin, Marepalli B Rao, and Ramachandran S Vasan. 2009. “Association between SNP Heterozygosity and Quantitative Traits in the Framingham Heart Study.” *Annals of Human Genetics* 73 (Pt 4):

465–73.

Grosen, Dorthe, Camilla Bille, Inge Petersen, Axel Skytthe, Jacob von Bornemann Hjelmborg, Jacob Krabbe Pedersen, Jeffrey Clark Murray, and Kaare Christensen. 2011. “Risk of Oral Clefts in Twins.” *Epidemiology* 22 (3): 313–19.

Hampe, Jochen, Andre Franke, Philip Rosenstiel, Andreas Till, Markus Teuber, Klaus Huse, Mario Albrecht, et al. 2007. “A Genome-Wide Association Scan of Nonsynonymous SNPs Identifies a Susceptibility Variant for Crohn Disease in ATG16L1.” *Nature Genetics* 39 (2): 207–11.

Hansen, T, A Skytthe, E Stenager, H C Petersen, H Brønnum-Hansen, and K O Kyvik. 2005. “Concordance for Multiple Sclerosis in Danish Twins: An Update of a Nationwide Study.” *Multiple Sclerosis* 11 (5): 504–10.

Hedden, Peter. 2003. “The Genes of the Green Revolution.” *Trends in Genetics: TIG* 19 (1): 5–9.

Hemani, Gibran, Sara Knott, and Chris Haley. 2013. “An Evolutionary Perspective on Epistasis and the Missing Heritability.” *PLoS Genetics* 9 (2): e1003295.

Hemani, Gibran, Konstantin Shakhbazov, Harm-Jan Westra, Tonu Esko, Anjali K Henders, Allan F McRae, Jian Yang, et al. 2014a. “Detection and Replication of Epistasis Influencing Transcription in Humans.” *Nature* advance online publication (February). doi:10.1038/nature13005.

———. 2014b. “Hemani et Al. Reply.” *Nature* 514 (7520): E5–6.

Hemani, Gibran, Athanasios Theocharidis, Wenhua Wei, and Chris Haley. 2011. “EpiGPU: Exhaustive Pairwise Epistasis Scans Parallelized on Consumer Level Graphics Cards.” *Bioinformatics* 27 (11): 1462–65.

Heng, Henry H Q. 2010. “Missing Heritability and Stochastic Genome Alterations.” *Nature*

Reviews. Genetics 11 (11): 813.

Hermisson, Joachim, and Günter P Wagner. 2004. “The Population Genetic Theory of Hidden Variation and Genetic Robustness.” *Genetics* 168 (4): 2271–84.

Hill, William G, Michael E Goddard, and Peter M Visscher. 2008. “Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits.” *PLoS Genetics* 4 (2): e1000008.

Holzinger, K J. 1929. “The Relative Effect of Nature and Nurture Influences on Twin Differences.” *Journal of Educational Psychology* 20 (4): 241–48.

Huerta-Sanchez, Emilia, and Rick Durrett. 2007. “Wagner’s Canalization Model.” *Theoretical Population Biology* 71 (2): 121–30.

“Identification of Risk Loci with Shared Effects on Five Major Psychiatric Disorders: A Genome-Wide Analysis.” 2013. *The Lancet* 381 (9875): 1371–79.

Jacquard, A. 1983. “Heritability: One Word, Three Concepts.” *Biometrics* 39 (2): 465–77.

Jacquemont, Sébastien, Bradley P Coe, Micha Hersch, Michael H Duyzend, Niklas Krumm, Sven Bergmann, Jacques S Beckmann, Jill A Rosenfeld, and Evan E Eichler. 2014. “A Higher Mutational Burden in Females Supports a ‘Female Protective Model’ in Neurodevelopmental Disorders.” *American Journal of Human Genetics* 94 (3): 415–25.

Jarosz, Daniel F, and Susan Lindquist. 2010. “Hsp90 and Environmental Stress Transform the Adaptive Value of Natural Genetic Variation.” *Science* 330 (6012): 1820–24.

Klei, Lambertus, Stephan J Sanders, Michael T Murtha, Vanessa Hus, Jennifer K Lowe, A J Willsey, Daniel Moreno-De-Luca, et al. 2012. “Common Genetic Variants, Acting Additively, Are a Major Source of Risk for Autism.” *Molecular Autism* 3 (1): 9.

Klempt, Martina, Birgit Rathkolb, Edith Fuchs, Martin Hrabé de Angelis, Eckhard Wolf, and Bernhard Aigner. 2006. “Genotype-Specific Environmental Impact on the Variance of Blood Values in Inbred and F1 Hybrid Mice.” *Mammalian Genome: Official Journal of the*

International Mammalian Genome Society 17 (2): 93–102.

Koga, Minori, Hiroki Ishiguro, Saori Yazaki, Yasue Horiuchi, Makoto Arai, Kazuhiro Niizato, Shuji Iritani, et al. 2009. “Involvement of SMARCA2/BRM in the SWI/SNF Chromatin-Remodeling Complex in Schizophrenia.” *Human Molecular Genetics* 18 (13): 2483–94.

Krumm, Niklas, Brian J O’Roak, Jay Shendure, and Evan E Eichler. 2014. “A de Novo Convergence of Autism Genetics and Molecular Neuroscience.” *Trends in Neurosciences* 37 (2): 95–105.

Kyvik, K O, A Green, and H Beck-Nielsen. 1995. “Concordance Rates of Insulin Dependent Diabetes Mellitus: A Population Based Study of Young Danish Twins.” *BMJ* 311 (7010): 913–17.

Lachowiec, Jennifer, Xia Shen, Christine Queitsch, and Örjan Carlborg. 2014. “Highly Epistatic Genetic Architecture of Root Length in Arabidopsis Thaliana.” *bioRxiv*, September, 008789.

Le Cunff, Yann, and Khashayar Pakdaman. 2012. “Phenotype-Genotype Relation in Wagner’s Canalization Model.” *Journal of Theoretical Biology* 314 (December): 69–83.

Lehner, Ben, Catriona Crombie, Julia Tischler, Angelo Fortunato, and Andrew G Fraser. 2006. “Systematic Mapping of Genetic Interactions in Caenorhabditis Elegans Identifies Common Modifiers of Diverse Signaling Pathways.” *Nature Genetics* 38 (8): 896–903.

Lempe, Janne, Sureshkumar Balasubramanian, Sridevi Sureshkumar, Anandita Singh, Markus Schmid, and Detlef Weigel. 2005. “Diversity of Flowering Responses in Wild Arabidopsis Thaliana Strains.” *PLoS Genetics* 1 (1): 109–18.

Levy, Sasha F, and Mark L Siegal. 2008. “Network Hubs Buffer Environmental Variation in Saccharomyces Cerevisiae.” *PLoS Biology* 6 (11): e264.

Li, Jing, Benjamin Horstman, and Yixuan Chen. 2011. “Detecting Epistatic Effects in Association Studies at a Genomic Level Based on an Ensemble Approach.” *Bioinformatics* 27 (13): i222–i229.

Lie, Hanne C, Leigh W Simmons, and Gillian Rhodes. 2009. “Does Genetic Diversity Predict Health in Humans?” *PloS One* 4 (7): e6391.

Lippert, Christoph, Jennifer Listgarten, Robert I Davidson, Scott Baxter, Hoifung Poong, Carl M Kadie, and David Heckerman. 2013. “An Exhaustive Epistatic SNP Association Analysis on Expanded Wellcome Trust Data.” *Scientific Reports* 3 (January). doi:10.1038/srep01099.

Lynch, Michael, and Bruce Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, Mass.: Sinauer.

Mackay, Trudy F C. 2014. “Epistasis and Quantitative Traits: Using Model Organisms to Study Gene-Gene Interactions.” *Nature Reviews. Genetics* 15 (1): 22–33.

Manolio, Teri A, Francis S Collins, Nancy J Cox, David B Goldstein, Lucia A Hindorff, David J Hunter, Mark I McCarthy, et al. 2009. “Finding the Missing Heritability of Complex Diseases.” *Nature* 461 (7265): 747–53.

McCarthy, S E, J Gillis, M Kramer, J Lihm, S Yoon, Y Berstein, M Mistry, et al. 2014. “De Novo Mutations in Schizophrenia Implicate Chromatin Remodeling and Support a Genetic Overlap with Autism and Intellectual Disability.” *Molecular Psychiatry* 19 (6): 652–58.

McClellan, Jon, and Mary-Claire King. 2010. “Genomic Analysis of Mental Illness: A Changing Landscape.” *JAMA: The Journal of the American Medical Association* 303 (24): 2523–24.

Mervis, Carolyn B, and Angela M Becerra. 2007. “Language and Communicative Development in Williams Syndrome.” *Mental Retardation and Developmental Disabilities*

Research Reviews 13 (1): 3–15.

Mitchell, Kevin J. 2012. “What Is Complex about Complex Disorders?” *Genome Biology* 13 (1): 237.

Mäki-Tanila, Asko, and William G Hill. 2014. “Influence of Gene Interaction on Complex Trait Variation with Multilocus Models.” *Genetics* 198 (1): 355–67.

Onstad, S, I Skre, S Torgersen, and E Kringlen. 1991. “Twin Concordance for DSM-III-R Schizophrenia.” *Acta Psychiatrica Scandinavica* 83 (5): 395–401.

O’Roak, Brian J, Pelagia Deriziotis, Choli Lee, Laura Vives, Jerrod J Schwartz, Santhosh Girirajan, Emre Karakoc, et al. 2011. “Exome Sequencing in Sporadic Autism Spectrum Disorders Identifies Severe de Novo Mutations.” *Nature Genetics* 43 (6): 585–89.

O’Roak, Brian J, Laura Vives, Wenqing Fu, Jarrett D Egertson, Ian B Stanaway, Ian G Phelps, Gemma Carvill, et al. 2012. “Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders.” *Science* 338 (6114): 1619–22.

O’Roak, Brian J, Laura Vives, Santhosh Girirajan, Emre Karakoc, Niklas Krumm, Bradley P Coe, Roie Levy, et al. 2012. “Sporadic Autism Exomes Reveal a Highly Interconnected Protein Network of de Novo Mutations.” *Nature* 485 (7397): 246–50.

Penrose, L S. 1969. “Effects of Additive Genes at Many Loci Compared with Those of a Set of Alleles at One Locus in Parent-Child and Sib Correlations.” *Annals of Human Genetics* 33 (1): 15–21.

Phelan, John P, and Steven N Austad. 1994. “Selecting Animal Models of Human Aging: Inbred Strains Often Exhibit Less Biological Uniformity Than F1 Hybrids.” *Journal of Gerontology* 49 (1): B1–B11.

Prescott, Natalie J, Sheila A Fisher, Andre Franke, Jochen Hampe, Clive M Onnie, Dianne

Soars, Richard Bagnall, et al. 2007. “A Nonsynonymous SNP in ATG16L1 Predisposes to Ileal Crohn’s Disease and Is Independent of CARD15 and IBD5.” *Gastroenterology* 132 (5): 1665–71.

Press, Maximilian O, Keisha D Carlson, and Christine Queitsch. 2014. “The Overdue Promise of Short Tandem Repeat Variation for Heritability.” *Trends in Genetics: TIG* 30 (11): 504–12.

Pritchard, J K, M Stephens, and P Donnelly. 2000. “Inference of Population Structure Using Multilocus Genotype Data.” *Genetics* 155 (2): 945–59.

Queitsch, Christine, Keisha D Carlson, and Santhosh Girirajan. 2012. “Lessons from Model Organisms: Phenotypic Robustness and Missing Heritability in Complex Disease.” *PLoS Genetics* 8 (11): e1003041.

Queitsch, Christine, Todd A Sangster, and Susan Lindquist. 2002. “Hsp90 as a Capacitor of Phenotypic Variation.” *Nature* 417 (6889): 618–24.

Ramanathan, Veerabhadran. 2009. “Newsmaker Interview: M. S. Swaminathan. A Guru of the Green Revolution Reflects on Borlaug’s Legacy. Interview by Pallava Bagla.” *Science* 326 (5951): 361.

Rendel, J M, J.M. 1962. “The Relationship between Gene and Phenotype.” *Journal of Theoretical Biology* 2: 296–308.

Richardson, Joshua B, Locke D Uppendahl, Maria K Traficante, Sasha F Levy, and Mark L Siegal. 2013. “Histone Variant HTZ1 Shows Extensive Epistasis With, but Does Not Increase Robustness To, New Mutations.” *PLoS Genetics* 9 (8): e1003733.

Rioux, John D, Ramnik J Xavier, Kent D Taylor, Mark S Silverberg, Philippe Goyette, Alan Huett, Todd Green, et al. 2007. “Genome-Wide Association Study Identifies New Susceptibility Loci for Crohn Disease and Implicates Autophagy in Disease

Pathogenesis." *Nature Genetics* 39 (5): 596–604.

Robertson, F W, and E C R Reeve. 1952. "Heterozygosity, Environmental Variation and Heterosis." *Nature* 170 (4320): 286.

Rohner, Nicolas, Dan F Jarosz, Johanna E Kowalko, Masato Yoshizawa, William R Jeffery, Richard L Borowsky, Susan Lindquist, and Clifford J Tabin. 2013. "Cryptic Variation in Morphological Evolution: HSP90 as a Capacitor for Loss of Eyes in Cavefish." *Science* 342 (6164): 1372–75.

Rosenfeld, Jill A, Bradley P Coe, Evan E Eichler, Howard Cuckle, and Lisa G Shaffer. 2013. "Estimates of Penetrance for Recurrent Pathogenic Copy-Number Variations." *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 15 (6): 478–81.

Rousseau, F, J Bonaventure, L Legeai-Mallet, A Pelet, J M Rozet, P Maroteaux, M Le Merrer, and A Munnich. 1994. "Mutations in the Gene Encoding Fibroblast Growth Factor Receptor-3 in Achondroplasia." *Nature* 371 (6494): 252–54.

Rutherford, Suzanne L. 2003. "Between Genotype and Phenotype: Protein Chaperones and Evolvability." *Nature Reviews. Genetics* 4 (4): 263–74.

Rutherford, Suzanne L, and Susan Lindquist. 1998. "Hsp90 as a Capacitor for Morphological Evolution." *Nature* 396 (6709): 336–42.

Sangster, Todd A, Neeraj Salathia, Hana N Lee, Etsuko Watanabe, Kurt Schellenberg, Keith Morneau, Hui Wang, Soledad Undurraga, Christine Queitsch, and Susan Lindquist. 2008. "HSP90-Buffered Genetic Variation Is Common in Arabidopsis Thaliana." *Proceedings of the National Academy of Sciences* 105 (8): 2969–74.

Sangster, Todd A, Neeraj Salathia, Soledad Undurraga, Ron Milo, Kurt Schellenberg, Susan Lindquist, and Christine Queitsch. 2008. "HSP90 Affects the Expression of Genetic Variation and Developmental Stability in Quantitative Traits." *Proceedings of the National*

Academy of Sciences of the United States of America 105 (8): 2963–68.

Satoh, Jun-Ichi, Megumi Nakanishi, Fumiko Koike, Hiroyuki Onoue, Toshimasa Aranami, Toshiyuki Yamamoto, Mitsuru Kawai, et al. 2006. “T Cell Gene Expression Profiling Identifies Distinct Subgroups of Japanese Multiple Sclerosis Patients.” *Journal of Neuroimmunology* 174 (1-2): 108–18.

Schönemann, P H. 1997. “On Models and Muddles of Heritability.” *Genetica* 99 (2-3): 97–108.

Sharma, Amitabh, Jörg Menche, C Chris Huang, Tatiana Ort, Xiaobo Zhou, Maksim Kitsak, Nidhi Sahni, et al. 2015. “A Disease Module in the Interactome Explains Disease Heterogeneity, Drug Response and Captures Novel Pathways and Genes in Asthma.” *Human Molecular Genetics*, January. doi:10.1093/hmg/ddv001.

Shiang, R, L M Thompson, Y Z Zhu, D M Church, T J Fielder, M Bocian, S T Winokur, and J J Wasmuth. 1994. “Mutations in the Transmembrane Domain of FGFR3 Cause the Most Common Genetic Form of Dwarfism, Achondroplasia.” *Cell* 78 (2): 335–42.

Slatkin, Montgomery. 2013. “A Method for Estimating the Effective Number of Loci Affecting a Quantitative Character.” *Theoretical Population Biology*, August. doi:10.1016/j.tpb.2013.08.002.

Song, Yun S, Fulton Wang, and Montgomery Slatkin. 2010. “General Epistatic Models of the Risk of Complex Diseases.” *Genetics* 186 (4): 1467–73.

Stahl, Eli A, Soumya Raychaudhuri, Elaine F Remmers, Gang Xie, Stephen Eyre, Brian P Thomson, Yonghong Li, et al. 2010. “Genome-Wide Association Study Meta-Analysis Identifies Seven New Rheumatoid Arthritis Risk Loci.” *Nature Genetics* 42 (6): 508–14.

Stefansson, Hreinn, Andreas Meyer-Lindenberg, Stacy Steinberg, Brynja Magnusdottir, Katrin Morgen, Sunna Arnarsdottir, Gyda Bjornsdottir, et al. 2014. “CNVs Conferring Risk of Autism or Schizophrenia Affect Cognition in Controls.” *Nature* 505 (7483): 361–66.

Steininger, Robert J, 3rd, Satwik Rajaram, Luc Girard, John D Minna, Lani F Wu, and Steven J Altschuler. 2014. “On Comparing Heterogeneity across Biomarkers.” *Cytometry. Part A: The Journal of the International Society for Analytical Cytology*, November. doi:10.1002/cyto.a.22599.

Thomas, Duncan. 2010. “Gene–environment-Wide Association Studies: Emerging Approaches.” *Nature Reviews. Genetics* 11 (4). Nature Publishing Group: 259–72.

Tsuang, Ming T, Jessica L Bar, William S Stone, and Stephen V Faraone. 2004. “Gene-Environment Interactions in Mental Disorders.” *World Psychiatry: Official Journal of the World Psychiatric Association* 3 (2): 73–83.

Visscher, Peter M, Matthew A Brown, Mark I McCarthy, and Jian Yang. 2012. “Five Years of GWAS Discovery.” *American Journal of Human Genetics* 90 (1): 7–24.

Visscher, Peter M, William G Hill, and Naomi R Wray. 2008. “Heritability in the Genomics Era–concepts and Misconceptions.” *Nature Reviews. Genetics* 9 (4): 255–66.

Wagner, Andreas. 2008. “Robustness and Evolvability: A Paradox Resolved.” *Proceedings. Biological Sciences / The Royal Society* 275 (1630): 91–100.

———. 2012. “The Role of Robustness in Phenotypic Adaptation and Innovation.” *Proceedings. Biological Sciences / The Royal Society* 279 (1732): 1249–58.

Walter, Henrik. 2013. “The Third Wave of Biological Psychiatry.” *Frontiers in Psychology* 4: 582.

Wilson, A J. 2008. “Why h2 Does Not Always Equal VA/VP?” *Journal of Evolutionary Biology* 21 (3). Blackwell Publishing Ltd: 647–50.

Wood, Andrew R, Marcus A Tuke, Mike A Nalls, Dena G Hernandez, Stefania Bandinelli, Andrew B Singleton, David Melzer, Luigi Ferrucci, Timothy M Frayling, and Michael N Weedon. 2014. “Another Explanation for Apparent Epistasis.” *Nature* 514 (7520): E3–E5.

Wray, Naomi R, Michael E Goddard, and Peter M Visscher. 2007. “Prediction of Individual Genetic Risk to Disease from Genome-Wide Association Studies.” *Genome Research* 17 (10): 1520–28.

Wray, Naomi R, Jian Yang, Ben J Hayes, Alkes L Price, Michael E Goddard, and Peter M Visscher. 2013. “Pitfalls of Predicting Complex Traits from SNPs.” *Nature Reviews. Genetics* 14 (7): 507–15.

Wright, Sewall. 1934. “The Results of Crosses between Inbred Strains of Guinea Pigs, Differing in Number of Digits.” *Genetics* 19 (6): 537–51.

Yamazaki, Keiko, Yoshihiro Onouchi, Masakazu Takazoe, Michiaki Kubo, Yusuke Nakamura, and Akira Hata. 2007. “Association Analysis of Genetic Variants in IL23R, ATG16L1 and 5p13.1 Loci with Crohn’s Disease in Japanese Patients.” *Journal of Human Genetics* 52 (7): 575–83.

Yang, Jian, Beben Benyamin, Brian P McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, et al. 2010. “Common SNPs Explain a Large Proportion of the Heritability for Human Height.” *Nature Genetics* 42 (7): 565–69.

Zaitlen, Noah, Peter Kraft, Nick Patterson, Bogdan Pasaniuc, Gaurav Bhatia, Samuela Pollack, and Alkes L Price. 2013. “Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits.” *PLoS Genetics* 9 (5): e1003520.

Zuk, Or, Eliana Hechter, Shamil R Sunyaev, and Eric S Lander. 2012. “The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (4): 1193–98.

Figure 1.

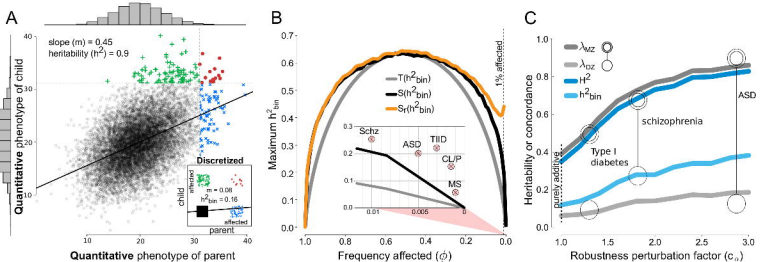

A
slope (m) = 0.45
heritability (h2) = 0.9
Quantitative phenotype of child
Quantitative phenotype of parent
Discretized
m = 0.08
h2bin = 0.16
affected
child
parent
B
Maximum h2bin
Frequency affected (ϕ)
T(h2bin)
S(h2bin)
Sr(h2bin)
1% affected
Schz
ASD
TIID
CL/P
MS
C
λMZ
λDZ
H2
h2bin
Heritability or concordance
Robustness perturbation factor (cu)
purely additive
schizophrenia
Type I diabetes
ASD

Figure 2.

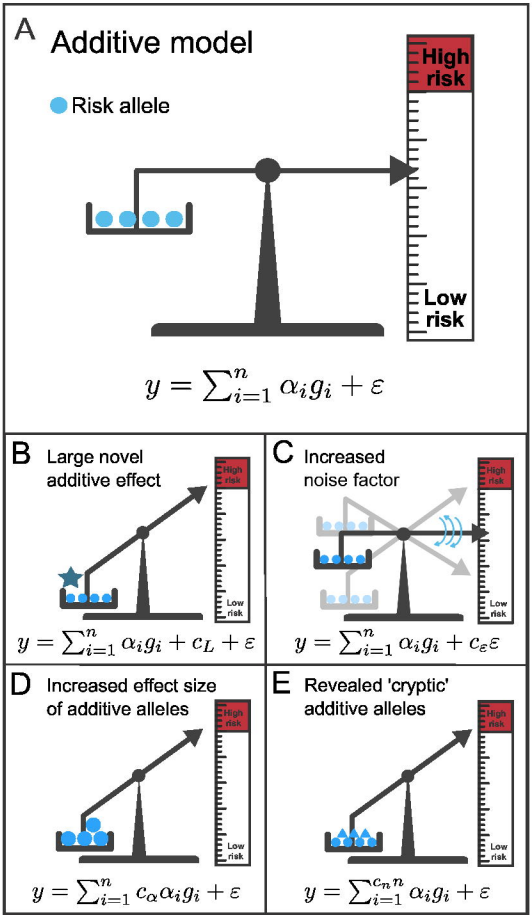

A
Additive model
Risk allele
High risk
Low risk
$y = \sum_{i=1}^{n} \alpha_i g_i + \varepsilon$
B
Large novel additive effect
$y = \sum_{i=1}^{n} \alpha_i g_i + c_L + \varepsilon$
C
Increased noise factor
$y = \sum_{i=1}^{n} \alpha_i g_i + c_\varepsilon \varepsilon$
D
Increased effect size of additive alleles
$y = \sum_{i=1}^{n} c_\alpha \alpha_i g_i + \varepsilon$
E
Revealed 'cryptic' additive alleles
$y = \sum_{i=1}^{c_n n} \alpha_i g_i + \varepsilon$

Figure 3.

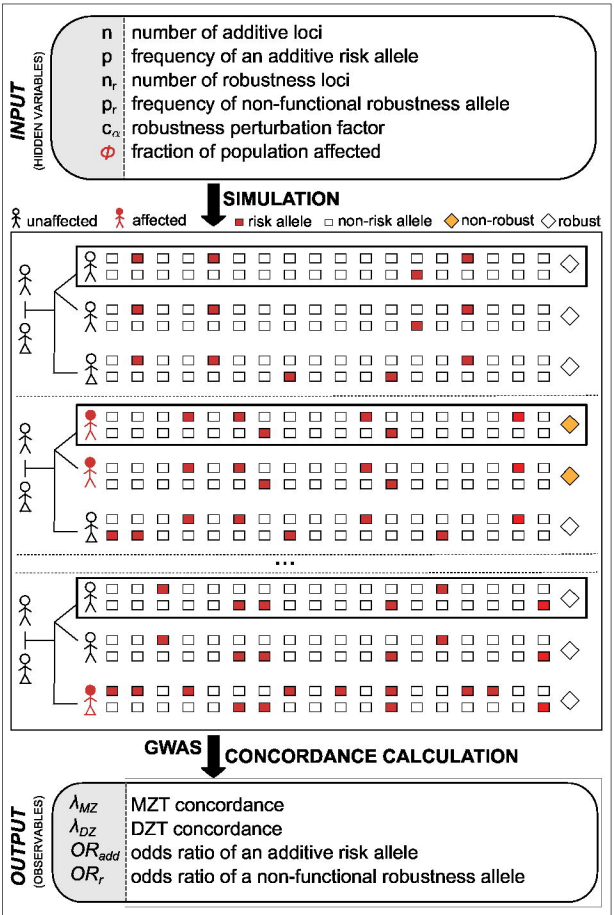
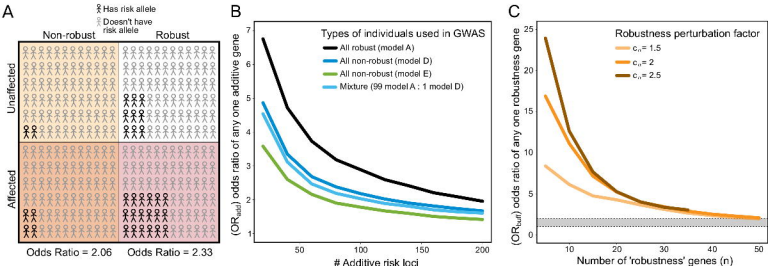

INPUT
(HIDDEN VARIABLES)
n number of additive loci
p frequency of an additive risk allele
n_r number of robustness loci
p_r frequency of non-functional robustness allele
c_α robustness perturbation factor
Φ fraction of population affected
SIMULATION
unaffected
affected
risk allele
non-risk allele
non-robust
robust
...
GWAS
CONCORDANCE CALCULATION
OUTPUT
(OBSERVABLES)
λ_{MZ} MZT concordance
λ_{DZ} DZT concordance
OR_{add} odds ratio of an additive risk allele
OR_r odds ratio of a non-functional robustness allele

Figure 4.


A
Has risk allele
Doesn't have risk allele
Non-robust
Robust
Unaffected
Affected
Odds Ratio = 2.06
Odds Ratio = 2.33
B
Types of individuals used in GWAS
All robust (model A)
All non-robust (model D)
All non-robust (model E)
Mixture (99 model A : 1 model D)
(OR_add) odds ratio of any one additive gene
Additive risk loci
C
Robustness perturbation factor
c_r = 1.5
c_r = 2
c_r = 2.5
(OR_out) odds ratio of any one robustness gene
Number of 'robustness' genes (n)