RESEARCH ARTICLE

Survival of the Curviest: Noise-Driven Selection for Synergistic Epistasis

Jon F. Wilkins[1☯], Peter T. McHale[2☯], Joshua Gervin[2], Arthur D. Lander[2]*

1 Ronin Institute, Montclair, New Jersey, United States of America, 2 Center for Complex Biological Systems & Department of Developmental and Cell Biology, University of California, Irvine, Irvine, California, United States of America

☯ These authors contributed equally to this work.
* adlander@uci.edu

Abstract

A major goal of human genetics is to elucidate the genetic architecture of human disease, with the goal of fueling improvements in diagnosis and the understanding of disease pathogenesis. The degree to which epistasis, or non-additive effects of risk alleles at different loci, accounts for common disease traits is hotly debated, in part because the conditions under which epistasis evolves are not well understood. Using both theory and evolutionary simulation, we show that the occurrence of common diseases (i.e. unfit phenotypes with frequencies on the order of 1%) can, under the right circumstances, be expected to be driven primarily by synergistic epistatic interactions. Conditions that are necessary, collectively, for this outcome include a strongly non-linear phenotypic landscape, strong (but not too strong) selection against the disease phenotype, and "noise" in the genotype-phenotype map that is both environmental (extrinsic, time-correlated) and developmental (intrinsic, uncorrelated) and, in both cases, neither too little nor too great. These results suggest ways in which geneticists might identify, a priori, those disease traits for which an "epistatic explanation" should be sought, and in the process better focus ongoing searches for risk alleles.

Author Summary

The contribution of epistasis, or non-additive effects of risk alleles at different loci, to complex traits is much debated among human geneticists. In this study we use modeling and simulation to identify when evolutionary forces should drive epistasis to become a major part of the explanation for such traits. We simulate populations evolving in the presence of "phenotypic noise", i.e. intrinsic and environmental variability in the relationship between genotype and phenotype, and focus specifically on traits that are substantially deleterious (e.g. fitness loss of at least 10%) and moderately common (population frequency between 0.1 and 10%). These criteria describe much "common", heritable human disease, and knowing the expected contribution of epistasis to such diseases could greatly assist efforts to use genome-wide genetic associations to identify their mechanistic underpinnings. Here we show that, provided that appropriate levels and kinds of phenotypic noise are present during evolution, genetic contributions to disease traits can be expected to be

strongly and synergistically epistatic, particularly among the most severely affected individuals in a population. By identifying scenarios most likely to select for synergistic epistasis, this study may help biologists identify when the extra effort is merited to search for combinatorial genetic interactions in human disease.

Introduction

Patterns of inheritance suggest that genetic variation plays a central role in the etiology of common, serious diseases (e.g., autism, schizophrenia, multiple sclerosis, diabetes). Yet genetic variants found by Genome Wide Association Studies (GWAS) typically explain only a small proportion of the heritability of such disorders [1]. Some have proposed that causative alleles elude discovery because their effect sizes are too small, or because they are not included among the variants that are normally ascertained. Another proposed reason is epistasis, i.e. the combinatorial action of variants at different genetic loci, such that they contribute significantly to disease only when they appear together in an individual [2–4]. Epistasis, also referred to as gene-gene interaction, can be exceedingly difficult to detect in GWAS—due to computational as well as statistical challenges [5,6]—and much debate exists about the extent to which efforts should be made to search for it.

Epistasis is common in the experimental genetics of model organisms, but one cannot simply extrapolate the behaviors of the large-effect alleles favored by experimental geneticists to the small-effect alleles thought to dominate the landscape of standing genetic variation (the fact that a continuous function can be approximated as linear over a small enough interval implies that, as selection drives the effect sizes of alleles toward zero, non-linear interactions—i.e. epistasis—should become insignificant). Although quantitative-trait-locus mapping in populations of model organisms has verified the existence of substantial epistasis among two [7–10] and even three or more genes at a time [11], the number of replicated examples of strong epistasis underlying human traits—particularly disease traits—is modest [12,13]. Whether this paucity of examples reflects the methodological difficulties (e.g. lower statistical power) inherent in working with human populations, or the fact that the major genetic causes of disease lie elsewhere (e.g. among rare, large-effect variants) is one of the most hotly debated questions in human genetics.

Progress toward addressing this question could be enhanced if investigators knew when and where to look for epistatic interactions in the human genome. One tactic is to use existing mechanistic knowledge—e.g. known biochemical or genetic pathways associated with the physiology disrupted by disease—to narrow the search, thereby improving statistical power [6]. The drawback to this approach is that, by working only outward from what we know, we sacrifice much of the ability to detect anything radically new. An alternative approach is suggested by the fact that epistasis, when present, must have evolved, i.e. come into being via the usual forces of mutation, drift, and natural selection. If we could determine the circumstances under which diseases with a strong epistatic genetic component do and do not evolve, we might be able to use that information as additional prior knowledge in genetic association studies.

The evolution of epistasis has previously been investigated by several groups, who have reached a variety of conclusions about it, not all of which are mutually compatible [14–18]. The approach in the present study differs from that prior work in that it is focused not on the typical or average behaviors of populations, nor on traits in general, but rather on genetic diseases—by which we mean traits that should be under strong negative selection—and populations in which disease incidence is in the range of many common, serious human diseases, i.e. between 0.1% and 10%. Using both theoretical analysis and evolutionary simulation, we find

that the extent to which epistasis explains a disease phenotype depends on the nature of the interaction between gene products (i.e. the biochemistry), the strength of selection on the trait, and especially on the nature and magnitude of stochastic and environmental fluctuations that influence gene function.

Results

Measures of Epistasis

To quantify epistasis is to capture the degree to which phenotypes cannot be accounted for by summing the phenotypic effects of variation at multiple genes, but one may approach this task in multiple ways. Among population geneticists, it is traditional to calculate the proportion of phenotypic variance, σ^2, that can be explained by an additive model, i.e. $h_{pop}^2 = \sigma_A^2 / \sigma^2$, where σ_A^2 represents the maximum variance that a linear model can produce, as determined by linear regression (Section 1 in S1 Text). The quantity h_{pop}^2 is termed the additive, or narrow-sense, heritability, and if non-genetic contributions to variance can be neglected or corrected for (a situation that we denote by an asterisk in what follows), $1 - h_{pop*}^2$ captures what is traditionally termed "statistical epistasis".

Alternatively, if one assumes that a particular phenotype or set of phenotypes may be treated as reference (i.e. "wildtype"), one may measure epistasis from the standpoint of an individual. For example, if $Z(g)$ represents the phenotype associated with a given genotype, g, then for a phenotype controlled by two loci, we may define

$$h_{ind*}(x, y | x_0, y_0) = \frac{[Z(x, y_0) - Z(x_0, y_0)] + [Z(x_0, y) - Z(x_0, y_0)]}{Z(x, y) - Z(x_0, y_0)}, \tag{1}$$

where $h_{ind*}(x,y|x_0,y_0)$ represents the proportion of phenotypic difference between genotype x,y and reference genotype x_0,y_0 that is due to additive effects (extension to a greater number of loci is straightforward). The asterisk in (Eq 1) reminds us that this definition neglects non-genetic contributions to the phenotype, generalization to which will be introduced later. In this case, $1 - h_{ind*}$ measures what is often termed "functional epistasis". Unlike statistical epistasis, functional epistasis may be further classified as synergistic ($h_{ind*} < 1$) or antagonistic ($h_{ind*} > 1$).

Although (Eq 1) is defined in terms of the genotype of a single individual, relative to a single reference, it is straightforward to generalize it to collections of "cases" and "controls", as we in fact do later. Thus, the primary difference between functional and statistical epistasis is not that one is population-centered and the other is individual-centered, but rather that functional epistasis focuses on the causes of variation from a pre-specified "wild" or "normal" state, whereas statistical epistasis focuses on the structure of variation within a population overall. From the standpoint of understanding the molecular causes of human disease, or predicting who will develop disease, functional epistasis is the more relevant quantity, as foreknowledge of a "normal" phenotype, or range of phenotypes, is implicit in the notion of disease. It is thus the evolution of functional epistasis that we are most concerned with here.

On the other hand, statistical epistasis is closely tied to the mechanics of evolution, because it is h_{pop*}^2 that effectively determines how natural selection acts on phenotypic variation [19]. This makes it relatively straightforward to develop insights about how statistical epistasis evolves, but such insights may be unhelpful with regard to functional epistasis, because the latter can exist in the absence, or near absence, of the former, e.g. [14,20–22]. How this can happen is shown in Fig 1.

Fig 1A depicts an arbitrary two-dimensional phenotypic landscape (i.e. allele values for each of two genes are on the independent axes, and phenotype value is on the dependent axis) and
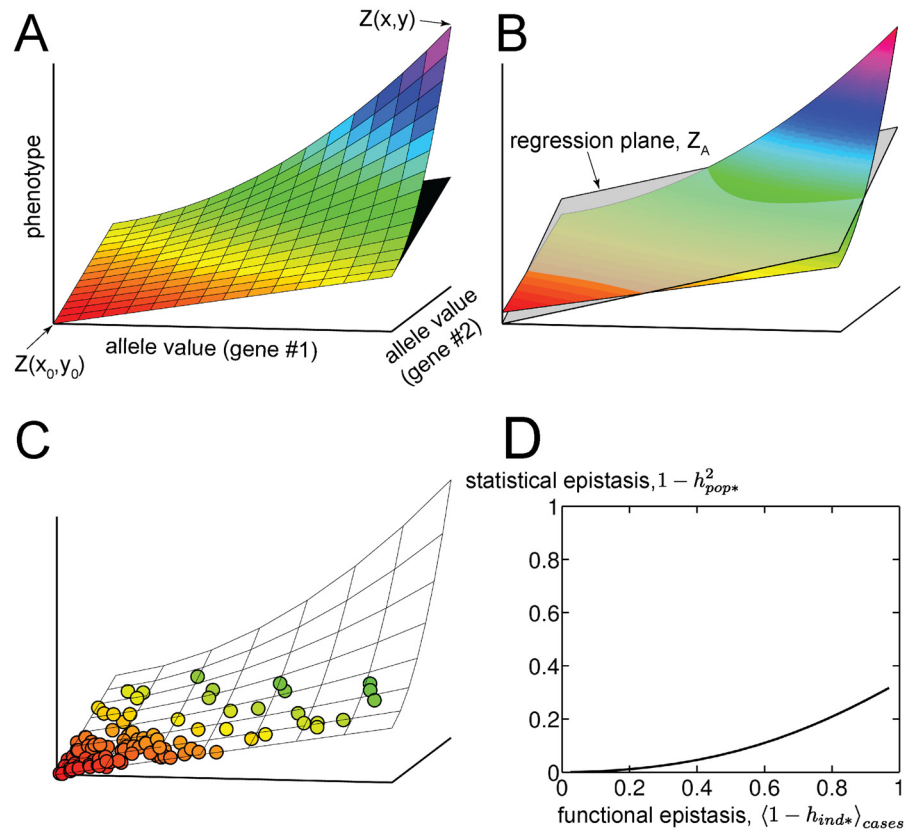
Fig 1. Epistasis at the individual versus population levels. (A) Functional epistasis. A phenotypic landscape (colored surface) is defined by Eq (S10) with ξ = 0.5 and n = 2. Allele values represent a quantitative measure of gene function given that the genotype is that allele. The difference between the phenotype value associated with two alleles ($Z(x,y)$), and the plane formed by the phenotype values of wildtype ($Z(x_0,y_0)$) and single-variant ($Z(x_0,y)$ and $Z(x,y_0)$) genotypes (shown in black), provides a local measure of epistasis. (B) Statistical epistasis. Linear regression determines a best-fit plane (translucent surface; see Eq (S2)) that depends upon the distribution of genotypes in the population. (C) A sample from the population used to perform the regression in B; individuals (circles) are colored according to their phenotype values. The population distribution is that defined by Eq (S11), with λ = 0.2 (D) Low statistical epistasis can exist in the presence of high functional epistasis. Here functional epistasis (evaluated with respect to the genotype (x_0,y_0); see panel A) has been averaged over all case individuals, defined as the 1% of the population with the greatest phenotype values. The parameter ξ was varied while the other parameters were held fixed at n = 1, λ = 0.2, C = 1 (see Eq. (S10) and (S11)).

illustrates how $h_{ind*}(x,y|x_0,y_0)$ simply reflects the nearness of $Z(x,y)$ (top-most point of colored landscape) to the plane formed by $Z(x_0,y_0)$, $Z(x_0,y)$ and $Z(x,y_0)$ (black surface visible directly below $Z(x,y)$). On the other hand, h_{pop*}^2 reflects how well the landscape is fit by a plane (translucent surface in [Fig 1B](#)), where the fit is weighted by the population distribution (a sample from one possible population distribution is shown in [Fig 1C](#)). In order for any sort of epistasis to occur, the landscape must be non-linear, but a large discrepancy between h_{ind*} and h_{pop*}^2 can occur at genotypes where the non-linearity is strong while the population density is low.

For example, it is relatively easy to construct scenarios in which nearly all of the phenotypic variation displayed by "diseased" individuals is explained by functional epistasis, even though there is little statistical epistasis in the population as a whole ([Fig 1D](#)). Even if diseased individuals represent as much as 50% of a study population (as in the artificially selected populations of "cases" and "controls" that are assembled for GWAS), statistical epistasis can still be well

below the level of functional epistasis of the cases, relative to the controls (S1A and S1B Fig; section 2 in S1 Text). In our evolutionary simulations, described below, we typically choose a phenotypic threshold between health and disease that assigns ~1% of the population to the latter category, consistent with the ascertainment level of many "common" human disorders. We thus focus on the questions of when, in the course of evolution, situations should arise in which functional epistasis is large among a subset of individuals of this magnitude.

Epistasis Is Selected against in the Simple LP Model

We began by considering a particularly simple model of non-additive gene-gene interaction—the Limiting Pathway (LP) model—that can be applied to a variety of biological processes [23]. In this model, a trait with value z depends on the rate-limiting value of a number of genetically controlled inputs, e.g., in the case when the number of inputs is two, $z = \text{Min}[x, y]$, where x and y represent the inputs from two genes (Fig 2A). These might represent the expression levels of two different polypeptides that combine to form a multi-subunit protein, the abundance of which is given by z. Alternatively, x and y could represent rates of synthesis of small molecules, such as enzyme cofactors, that catalyze a process occurring at rate z. For systems involving more than two inputs, we have $z = \text{Min}[x_1, x_2, x_3 . . .]$, but for simplicity we will focus here on the two-component model.

We assume there is an optimal phenotype value, z_{opt}, and that the trait quantified by z is under stabilizing selection, so that fitness is reduced when z falls above or below z_{opt}. Thus, at evolutionary equilibrium, we expect the wild-type levels of the gene products, which we will call x_0 and y_0, to be such that $\text{Min}[x_0, y_0] = z_{opt}$. The representation of this condition on the phenotypic landscape, i.e. the intersection of the plane $z = z_{opt}$ with the function $z = \text{Min}[x, y]$, defines two perpendicular "arms", meeting at a "corner" (green dashed lines in Fig 2A).

Only when populations reside near this corner, *i.e.*, $x_0 \approx y_0$, can significant epistasis occur. For example, under such conditions mutations that increase x or y may have little phenotypic effect individually, but a large effect in combination (synergistic epistasis). In contrast, for a population residing on one of the arms—e.g., $x_0 >> y_0$, corresponding to the horizontal arm in Fig 2A—mutations that increase or modestly decrease x will have no phenotypic effect; mutations that affect y will alter z proportionately; and the combined effect of both mutations will be no different from the sum of their individual effects.

For any pair of wild-type input values (x_0, y_0) and mutation effect sizes $(\Delta x, \Delta y)$, we may quantify the magnitude of functional epistasis as $1 - h_{ind^*}(x_0 + \Delta x_0, y_0 + \Delta y_0 \mid x_0, y_0)$, according to (Eq 1), as described above. For a given mutation effect ($\Delta x = \Delta y > 0$), we repeat this calculation for different combinations of wild-type values x_0 and y_0, to produce an epistasis map (Fig 2B). Asking whether epistasis will tend to evolve in the LP model thus amounts to asking whether input values in natural populations will evolve towards and remain at the corner in Fig 2A. Re-plotting Fig 2A in terms of fitness, which we take to be the following generic function of phenotype value

$$w(z) = \exp\left(-s\left(\ln\frac{z}{z_{opt}}\right)^2\right), \tag{2}$$

(where the strength of stabilizing selection is quantified by s) yields an L-shaped ridge on which any point along the ridge is equally and maximally fit (Fig 2C). Given this fitness landscape, our naïve expectation might be that, under selection and drift, populations should follow a random walk along the ridge, with individual input values equally likely to be clustered around any ridge location.
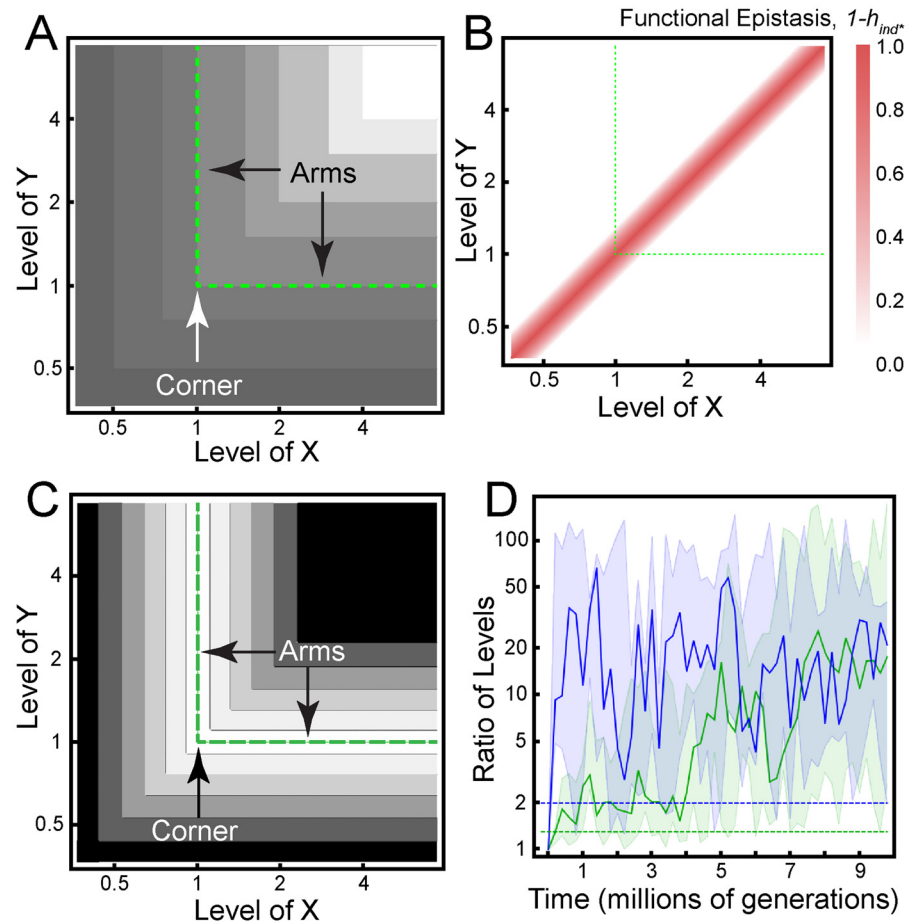
Fig 2. Epistasis does not evolve in the simple LP model. (A) Phenotype value as a function of input values in the simple LP model, $z = \text{Min}[x,y]$. Lighter shading represents higher phenotype values. (B) At each combination of x and y, we calculated the contribution of epistasis to the change in phenotype resulting from a 30% increase in each input value. Near the diagonal, the "sign" of local functional epistasis depends on whether mutations increase or decrease x and/or y. Mutations that increase (decrease) both x and y result in synergistic (antagonistic) epistasis. The nature of epistasis between a mutation that increases one input value and a mutation that decreases the other will depend on the precise values of x and y. (C) Fitness landscape given by ([Eq 2](#)), with the optimal value of z being $z_{opt} = 1$ and the strength of selection being $s = 1$. Lighter shading represents higher fitness. (D) Simulated time course of a population evolving along the optimal-fitness ridge indicated by the green dashed lines in C. The population-mean input values (i.e. average of individual (x,y) values weighted by their corresponding frequencies in the population) were calculated every 200,000 generations and their ratio (y/x if $y>x$; x/y if $x>y$) recorded on the vertical axis, which indicates how far the population center of mass is from the corner of the ridge. The green band and curve indicate the minimum, maximum, and median values from five independent simulations with an average mutational effect size of $\sigma_\mu = 0.1$. The blue band and curve correspond to a larger mutational effect size ($\sigma_\mu = 1.0$). The region around the corner over which we expect to observe functional epistasis is limited by the effect size of a typical mutation (indicated by the horizontal dashed lines drawn at a distance of σ_μ from the corner). Initially, all individuals in the population lie at the corner ($x = y = 1$).

However, evolutionary simulations (see [Methods](#)) show that selection drives populations away from the corner and out onto the arms of the ridge ([Fig 2D](#)). To understand why, note that at "arm" locations where $y \ll x$, mutations affecting x will be selectively neutral. But if the population is close to the corner, such that $y < x < y + \sigma_\mu$, where σ_μ is the typical mutational step size, mutations that reduce x will be deleterious. Because of the increased frequency of deleterious mutations near the corner, natural selection disfavors genotypes with input values

within σ_μ of the corner (compare shaded regions at long times with horizontal dashed lines in Fig 2D). It follows that the population is least likely to be found precisely where epistasis is strongest.

A More Biologically Realistic LP Model

Is the evolutionary instability of epistasis in the LP model something common to all genotype-phenotype relationships in which strong synergistic epistasis can arise? Or is the LP model, as currently formulated, insufficiently general? To approach this question, we explored two modifications of the LP model.

First, we generalized the phenotype function with the aid of parameter k, which allows us to vary the strength of interaction between pathways:

$$z = \frac{1}{2}\left(k + x + y - \sqrt{(k + x + y)^2 - 4xy}\right) \tag{3}$$

When $k = 0$, (Eq 3) reduces to the simplified LP model, *i.e.* $z = \text{Min}[x,y]$. If x and y were to represent the concentrations of two reactants, and z the concentration of their product, then k would represent the corresponding biochemical dissociation constant. In this interpretation, if we choose our units so that $z_{opt} = 1$, $k >> 1$ would correspond to weak binding while $k << 1$ would be tight binding. We may also interpret x and y in (Eq 3) as rates of synthesis of gene products, though the physiological meaning of k would then be different, incorporating binding and rates of degradation. As shown in Fig 3A and 3B, the effect of increasing k is to make the corner in the fitness landscape more round. This increases (for any given mutational step size) the range of values of x and y over which significant epistasis will occur, but also decreases the maximum potential magnitude of that epistasis.

Second, we added phenotypic noise into the model. By noise we mean processes that add randomness into the genotype-phenotype relationship. For purposes of this analysis, we divide noise into two categories: developmental and environmental.

By developmental noise, we mean processes that affect each individual in the population independently and are not substantially correlated between parents and offspring (Fig 3C). An example of developmental noise would be idiosyncratic variation in the biochemical processes underlying development, such that individuals with identical genotypes do not necessarily display identical phenotypes. Other factors that could contribute to developmental noise by this definition would include micro-environmental effects that vary from individual to individual and genetic variation at loci not explicitly incorporated into the model (assuming high recombination).

By environmental noise, we mean processes that have a coordinated effect on all individuals in a population, and may persist from one generation to the next (Fig 3D). Processes like climate change, dietary change, and cultural change, which have the potential to fluctuate on time scales of many generations, would fall into this category. Although it is common to model such processes as perturbations to the fitness (i.e., "fluctuating selection", e.g., [24]), we model these processes as perturbations to phenotype, in order to facilitate comparison between the two classes of noise.

Accordingly, we take noise generally to be a process that transforms an individual's *nominal* input values (the values of x and y dictated by the individual's genotype; white circles in Fig 3C and 3D; S2A and S2C Fig) into an *effective* set of input values (red circles in Fig 3C and 3D; S2B and S2D Fig) that produce, via (Eq 3), an altered phenotype. The effective input values can be thought of as the set of values that would produce the observed phenotype in the absence of any noise.
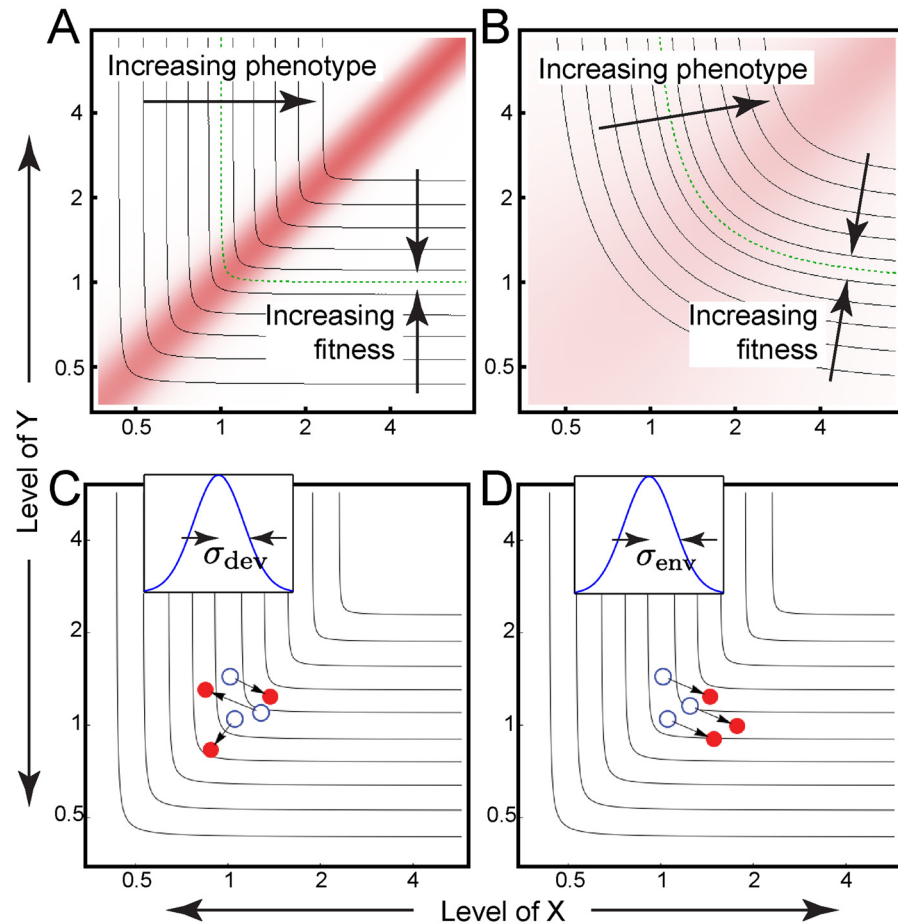
Fig 3. The generalized LP model. (A, B) Epistasis landscape (red shading) and common contours of the phenotype and fitness landscapes (black lines) for k = 0.005 (A) and 0.5 (B). (C, D) Schematic illustrating how phenotypic noise can be viewed as random perturbations of the input values, transforming nominal values (white) into effective values (red). The magnitudes of the perturbations are drawn from a Gaussian distribution defined on the natural logarithm of gene levels (insets). This sampling process is repeated for each individual independently (developmental noise; C) or performed once for all individuals (environmental noise; D); see Methods.

Motivated by the fact that cell-to-cell variation in the levels of gene products is often found to fit log-normal distributions [25,26], we implement developmental noise in x, say, via the transformation

$$\ln x_e = \ln x + \Delta x \tag{4}$$

where Δx is drawn from a normal distribution with mean zero and variance σ_{dev}^2 (Fig 3C; inset). Perturbations to y are performed independently using the analogous transformation.

With environmental noise, the same perturbation affects the x- and y-values of all individuals in the population (Fig 3D; inset). Although that perturbation is also drawn randomly (this time from a zero-mean Gaussian with standard deviation σ_{env}), it is not re-drawn at random every generation, as environmental effects may vary on slow time scales. Details of how this is modeled are provided in the Methods section, but for all results presented in the main text the Δx and Δy values for environmental noise have an autocorrelation time of 27 generations.

Noise Stabilizes the Corner

How should we expect phenotypic noise to affect population dynamics on the generalized LP landscape? So long as the autocorrelation time of phenotypic noise is short relative to the time-scale on which natural selection drives genetic change, we can approximate the effect of natural selection on a particular combination of input values by averaging over their different fitness realizations due to noise. The resulting "effective fitness" measures how robust the phenotype (and nominal fitness) of that particular set of input values is to random variation of their values.

When we compute this effective fitness landscape, we find a narrow peak at the corner for $k = 0$ (Fig 4A) that develops into a broad peak at $k = 0.1$ (Fig 4B). Though a peak at the corner is intuitively expected to localize the population there, it is less clear how the *shape* of the peak affects localization. Since the motion of the population is driven by fitness differences, a narrow peak (and a steep fitness gradient) should exert a strong force driving the population towards the corner. Yet that force can only be felt in the immediate vicinity of the corner, potentially allowing populations located elsewhere to drift away from the corner. On the other hand, though a broad peak exerts a weaker force (due to its shallower fitness gradient), its range of influence is greater, potentially drawing populations to the corner that would have otherwise escaped it.

To find out which landscape best stabilizes the corner at steady state, we mapped the evolution of a population on the 2D effective fitness landscape to a 1D random walk problem (Section 3 of S1 Text). The random-walk theory predicts that localization at the corner depends monotonically on k, with the greatest degree of corner localization expected at the smallest values of k (Fig 4D)—precisely the conditions under which epistasis is strongest (Fig 3A). Assuming that k is low enough, we next asked how corner localization depends on the noise level, by which we mean the width of the Gaussian from which the phenotypic perturbation is drawn (σ_{dev} or σ_{env}; Fig 3C and 3D). When the noise level drops below 0.1, the random walk is effectively unbiased, with the effective fitness profile being approximately flat (Fig 4E) and the probability of finding the walker at least a given distance from the corner falling linearly with distance (Fig 4F).

The random-walk theory, though providing significant insight, is an approximation that is expected to hold only when the product of the population size and the mutation rate is small, i.e. $N\mu << 1$. To address more general conditions, we examined the population dynamics using evolutionary simulations. Time courses plotted from individual simulations show that both developmental and environmental noise are effective at localizing the population in proximity of the corner (where epistasis can arise) when the interaction parameter, k, is small enough (Fig 5A and 5B; see also S2E and S2G Fig). To summarize the degree of corner localization under a given set of parameter values, we sampled simulations at intervals of $2N$ generations (where N is population size) and computed the fraction of simulations where the population is at least a given distance from the corner (Fig 5C and 5D). For both developmental and environmental noise, simulated data confirm the general trend predicted by the random-walk theory: the sharper the corner (and the greater the potential for strong epistasis), the more effective phenotypic noise is at localizing the population there (see also S4 Fig).

We repeated the evolutionary simulations for a range of noise levels, either pure developmental or pure environmental. Fig 5E shows that the dependence of corner localization upon developmental noise is remarkably similar to that predicted by the random-walk theory (Fig 4F). However, when the phenotypic noise is environmental, the results of theory and simulation diverge. Evolutionary simulations show that corner localization depends non-monotonically on environmental noise level such that only intermediate noise levels ($\sigma_{env} \approx 0.1$) hold the population near the corner (Fig 5F). In contrast, the theoretical dependence on noise level was always monotonic (Fig 4F).
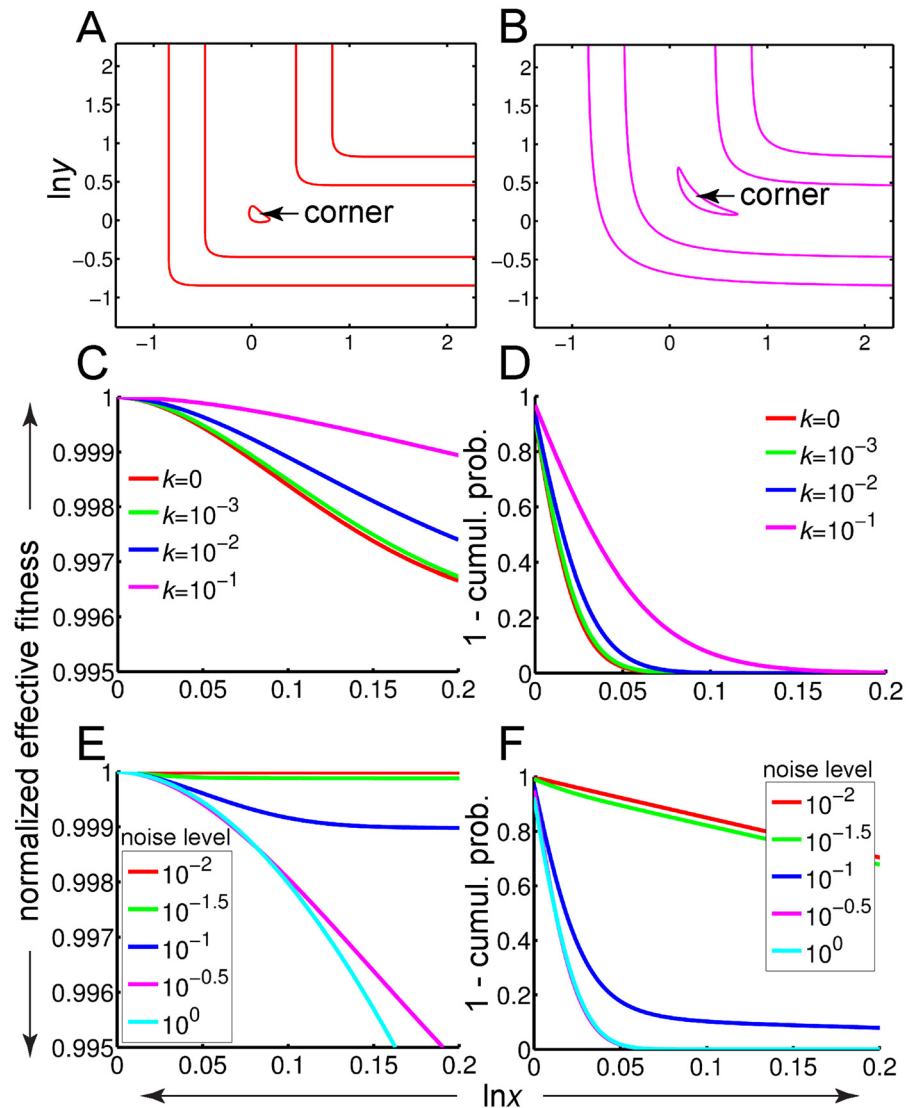
Fig 4. Phenotypic noise stabilizes the corner in the random-walk approximation. (A, B) Contours of the effective fitness landscape, $w_e(x,y)$, for $k = 0$ (A) and $k = 0.1$ (B). Effective fitness resulting from phenotypic noise is computed at any given position by averaging neighboring fitness values using a rotationally symmetric normal distribution (on log-transformed input values) with a standard deviation ("noise level") of 0.2. Contour lines correspond to $w_e = 0.5, 0.8, 0.99$. (C) Normalized effective fitness profiles, $w_e(x')/w_e(0)$, along the pseudo-horizontal arm of the optimal fitness ridge, projected onto the x-axis and shifted horizontally so as to align the corners. $x' = \ln x$. (D) Probabilities that the population is found at least the indicated distances from the corner. (E, F) Same as panels C and D, respectively, except now the noise level is changed while the interaction constant is kept fixed at $k = 10^{-4}$. Throughout, population size is $N = 5000$, optimal phenotypic value is $z_{opt} = 1$ and selection factor is $s = 1$.

The accuracy of random-walk theory in predicting the role of developmental but not environmental noise in simulations reflects the fact that the theory takes as input an *effective* fitness landscape representing the average effect of phenotypic noise. While such averaging is expected to faithfully capture pure developmental noise, which is independently realized for each individual and at each generation (Fig 3C), this approximation does not account for the correlations (among individuals and across generations) intrinsic to our implementation of environmental noise (Fig 3D).
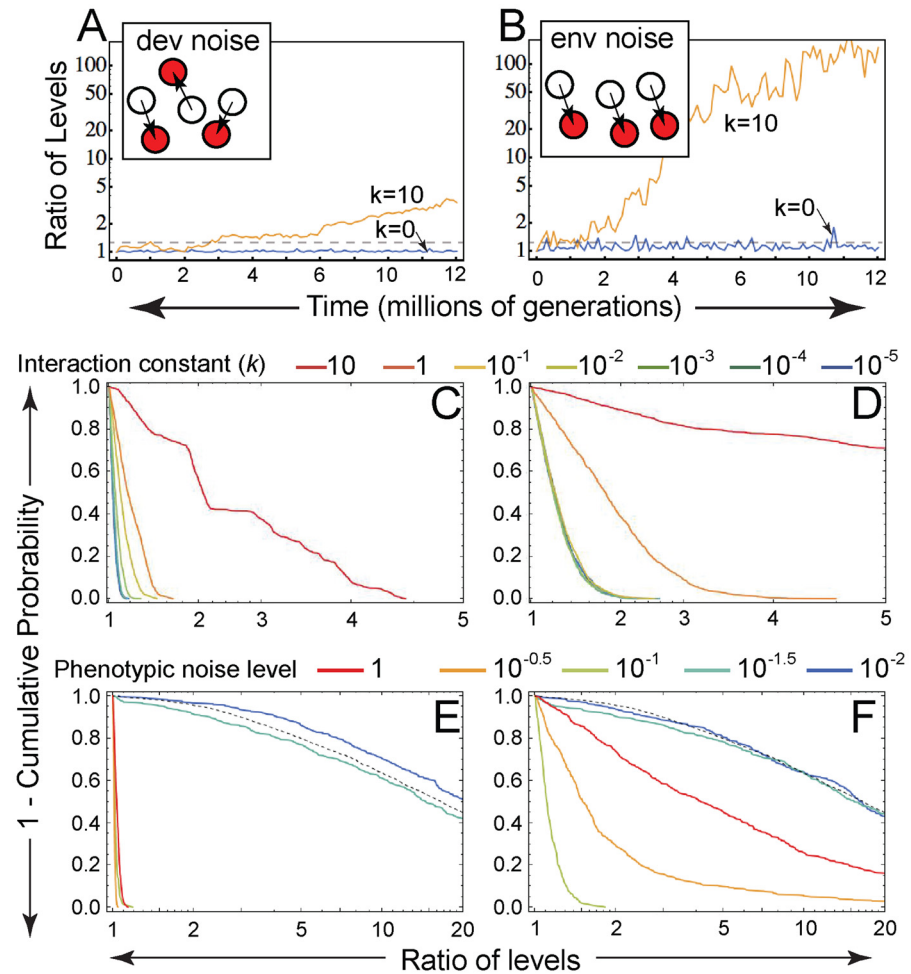
Fig 5. How the character of phenotypic noise affects corner stabilization in simulations. (A, B) The full evolutionary model (see Methods) was used to simulate time courses of the population center of mass (solid lines) for two different values of the interaction constant, k, subject to either pure developmental noise (A) or pure environmental noise (B). The proximity of the corner within which epistasis is expected is indicated by the horizontal dashed lines drawn at a distance of $\sigma_\mu = 0.1$ from the corner. (C–F) The probability that the population is at least a given distance from the corner for various values of k (C, D), and various levels of developmental and environmental noise, σ_{dev} and σ_{env}, respectively (E, F). In A and C, the phenotypic noise level is $\sigma_{dev} = 0.1$; in B and D it is $\sigma_{env} = 0.1$. In E and F, $k = 10^{-4}$, and the black dashed line is the limiting case of no phenotypic noise ($\sigma_{dev} = \sigma_{env} = 0$). Mutational step size in C–F was $\sigma_\mu = 1$. Throughout, data in the left and right columns correspond to purely developmental and purely environmental phenotypic noise, respectively.

doi:10.1371/journal.pgen.1006003.g005

In summary, both theory and simulation show that phenotypic noise drives populations to the corner of the generalized LP fitness landscape, with the greatest degree of corner localization expected at the smallest values of k—the same condition under which the genotype-phenotype map is most nonlinear (Fig 3A). Though such nonlinearity is clearly a necessary condition for generating epistasis in individuals of the population, it is certainly not sufficient. To shed light on human disease biology, we next engaged the full power of our simulations to determine the conditions under which individuals affected by disease exhibit epistasis relative to healthy controls, and the frequencies at which such individuals may arise.

Expectations for Epistasis in Case-Control Studies

We modeled case-control studies of common disease (prevalence $\approx 1\%$) by sampling individuals with extreme and typical phenotype values (Fig 6A; S2 Fig), which we denoted "cases" and "controls", respectively. For each case individual, we partitioned the difference between its phenotype value, z_{case}, and the phenotype associated with the median input values of the control samples, $z_{controls}$, into a fraction corresponding to the difference in nominal input values,

$$H_{ind} = \frac{z_H - z_{controls}}{z_{case} - z_{controls}}, \tag{5}$$

which we refer to as the "heritability", and a non-heritable fraction, 1-H_{ind}. The quantity z_H represents the heritable phenotype of the case individual, which we approximated by its phenotype value before *developmental* noise was added (Fig 6B). In contrast, *environmental* noise is not expected to affect phenotype heritability assessed either within a generation (since genetically identical individuals have the same phenotype even after the addition of environmental noise) or between two consecutive generations (since environmental noise persists over many generations). Note that H_{ind}, closely resembles the concept of broad-sense heritability as used by human geneticists, but is defined here at the level of a single individual, rather than the population as a whole. We can naturally partition H_{ind} into an additive fraction,

$$h_{ind} = \frac{(z_x - z_{controls}) + (z_y - z_{controls})}{z_{case} - z_{controls}}, \tag{6}$$

and an "epistatic fraction", H_{ind}—h_{ind}, where z_x and z_y represent the phenotype values of the single mutants defined by the heritable input values of the case individual (Fig 6B).

Epistatic fraction is a generalized measure of functional epistasis that is valid in the presence of phenotypic noise. To see this, note that, in the absence of noise, H_{ind} reduces to 1 and h_{ind} reduces to $h_{ind^*}(x,y \mid x_0, y_0)$, where h_{ind^*} is defined by (Eq 1) and the input values (x, y, etc) are determined by the relations $Z(x,y) = z_H = z_{case}$ and $Z(x_0,y_0) = z_{controls}$. Thus H_{ind}—h_{ind} reduces to 1—h_{ind^*}, which is the definition of functional epistasis presented earlier. The no-noise limit is instructive for another reason: it turns out that h_{ind^*} evaluated at one population standard deviation relative to the median control sample directly determines the narrow-sense heritability used in human genetics, $h^2_{pop^*}$ (S1C Fig; see also Section 1 in S1 Text).

We measured the epistatic fraction of case-control phenotype difference, H_{ind}—h_{ind}, and the fitness (given by (Eq 2), normalized by the mean fitness of the control samples) for each case individual and for each time point sampled during the course of our evolutionary simulations. Fig 6C (data aggregated over all time points) and 6D (data stratified by time point) both show that, though the vast majority of individual case phenotypes are largely additive ($H_{ind} \approx h_{ind}$), significantly unfit case individuals (i.e. fitness < 0.9) are *always* predominantly epistatic (H_{ind}—$h_{ind} > 0.5$) with respect to controls. Put another way, although severe disease may arise infrequently in these simulations, when it does, the explanation for it is generally epistatic.

Because these evolutionary simulations keep track of individual mutations across generations, we can also calculate the frequency, in simulated populations, of the individual alleles that account for the epistatic interactions that give rise to unfit cases. As S5 Fig. shows, alleles involved in producing exceedingly unfit individuals (fitness < 0.75) tend to be relatively rare (frequency between 0.2% and 0.5%), whereas in somewhat less severely unfit cases (fitness between 0.75 and 0.9), causal alleles may display frequencies $>1\%$, in the range of frequencies associated with minor alleles that can be followed in GWAS.

Fig 6E shows how the epistatic fraction, H_{ind}-h_{ind}, averaged over all substantially unfit cases (those with relative fitness < 0.9), varies with changing levels of phenotypic noise. In
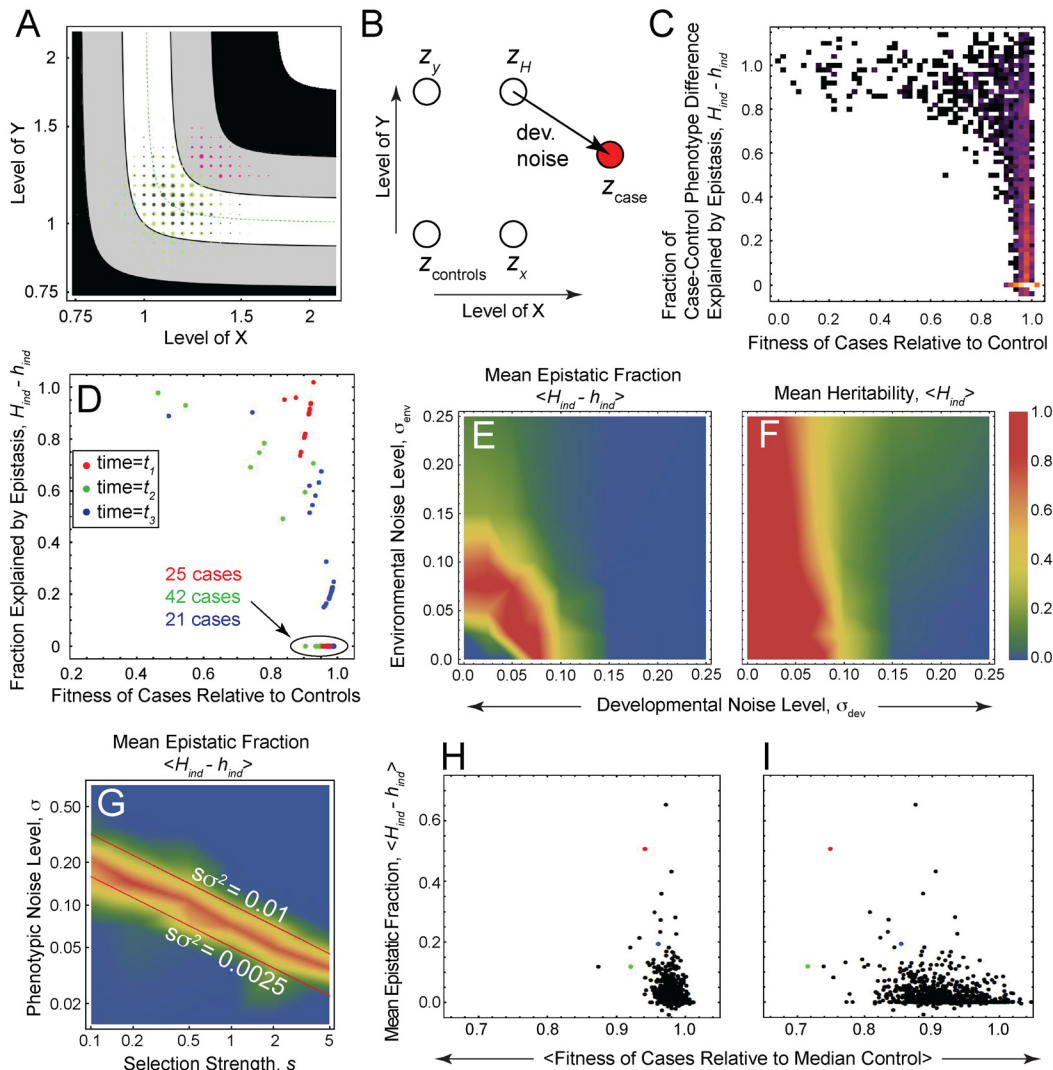
Fig 6. When to expect epistasis in case-control studies. (A) Distribution of *effective* input values among all individuals (green) in a simulated population and conditioned upon individuals being "cases" (pink) or "controls" (dark blue). Cases consist of the 1% of the population with the highest phenotype values; 500 controls are sampled from the 50% of the population whose phenotype values lie between the 25th and 75th percentiles. The area of each circle is proportional to the number of individuals whose effective input values fall within a square bin at the circle's location. Other parameter values are $z_{opt} = s = 1$; $N = 5000$; $k = 0.01$; $\sigma_\mu = 0.1$; $\sigma_{dev} = 0.1$; $\sigma_{env} = 0$. (B) Illustration showing how phenotype differences between cases and controls are partitioned into heritable (additive plus epistatic) and non-heritable components (see main text). $z_{controls}$ represents the phenotype of a hypothetical control whose position is given by the median of the effective input values (i.e. including developmental noise) of all actual controls. (C) Joint distribution of the epistatic fraction, $H_{ind} - h_{ind}$, and fitness (relative to mean control fitness) over all cases (top 1% of phenotypes) and all sampled time points. The color of each pixel reflects the number of cases with the corresponding fitness and epistasis values. Parameter values are $k = 0$ and $\sigma_{dev} = \sigma_{env} = 0.05$. (D) Joint distribution of epistasis and fitness of cases (top 0.5% of phenotypes) for each of three time points, t_1, t_2, t_3, distinguished by color. Parameter values are $\sigma_{dev} = \sigma_{env} = 0.05$. (E—G) Heat maps with intensity representing epistatic fraction, $H_{ind} - h_{ind}$ (E, G) and heritability, H_{ind} (F) averaged over all cases with relative fitness < 0.9 and all sampled time points. In G, the y-axis indicates the total noise level $\sigma = \sqrt{\sigma_{dev}^2 + \sigma_{env}^2}$, with $\sigma_{dev} = \sigma_{env}$, and $\sigma_\mu = 1$. (H) Distribution over all sampled time points of epistasis and fitness (each averaged over all cases at the corresponding time point). Colored data points correspond to the three time points sampled in panel D. (I) Same as H but, at each sampled time point, case phenotype and case fitness were assessed immediately after an abrupt 5-fold increase in selection strength from $s = 1$ to $s = 5$. The corresponding fold change in relative fitness of each case is approximately $w^{(s'/s)-1}$ (which follows from ([Eq 2])), where w is the relative case fitness prior to the selection shift. See also [S12 Fig]. Parameter values in H and I were $k = 0$; $\sigma_\mu = 1$; $\sigma_{dev} = \sigma_{env} = 0.05$.

doi:10.1371/journal.pgen.1006003.g006

the absence of developmental noise (vertical axis of Fig 6E), strong, synergistic epistasis ($<H_{ind}-h_{ind}> > 0.5$) is observed only when environmental noise is of intermediate strength, because only then is the population localized to the corner (Fig 5F). More surprising, at first sight, is the fact that, in the absence of environmental noise (horizontal axis of Fig 6E), developmental noise must also be of intermediate strength, and is even more constrained than environmental noise (compare the extent of the red region along the two cardinal axes in Fig 6E). This constraint on developmental noise level originates in a trade-off between its effects on corner localization and (total) heritability, H_{ind}. In the absence of environmental noise, modest developmental noise localizes the population to the corner (Fig 5E), favoring epistasis ($\sigma_{dev}<0.075$; Fig 6E); on the other hand, large-amplitude developmental noise kills heritability (Fig 6F; S2A Fig), and with it, epistasis ($\sigma_{dev}>0.075$; Fig 6E), which is bounded above by heritability (recall that heritability, H_{ind}, is partitioned into an additive portion, h_{ind}, and an epistatic portion, $H_{ind}-h_{ind}$). Thus, there is a range of developmental noise levels ($\sigma_{dev}\approx0.05–0.10$; Fig 6E) capable of localizing the population near the corner while preserving heritability.

How does the degree of epistasis among unfit (fitness <0.9) cases depend upon the other key ingredient of the extended LP model—the interaction parameter, k? While decreasing k below 0.01 does not affect corner stability (S4 Fig), it does increase the mean epistatic fraction, $<H_{ind}-h_{ind}>$ significantly (S6 Fig; also compare Fig 3A with 3B). Thus, by making the LP landscape as nonlinear as possible, by sending k to zero, one maximizes the strength of epistasis one expects to observe in unfit case individuals.

The fact that the relative fitness of a case individual is such a strong predictor of its probability of being epistatic (Fig 6C and 6D and S7 Fig) prompted us to ask how epistasis is affected by the degree to which a trait is under selection, as measured by the selection factor, s, in (Eq 2). Fig 6G shows that the trait of interest must be under a certain amount of selection before we may expect strong epistasis among severely affected cases. That level of selection varies inversely with the level of phenotypic noise, $\sigma = \sqrt{\sigma_{dev}^2 + \sigma_{env}^2}$, as indicated by the red band in Fig 6G. In fact, this figure shows that epistasis depends primarily on the product $s\sigma^2$ (see also Section 4 in SI), so that increasing the strength of selection is equivalent to increasing the noise variance. Epistasis is greatest for intermediate values of $s\sigma^2$ (in the vicinity of 0.005), as previously seen in Fig 6E, which varies σ while holding the selection factor fixed at $s = 1$.

In summary, evolutionary simulations indicate that strong synergistic epistasis can underlie disease that arises in populations evolving on a generalized LP landscape. Neither disease nor epistasis tend to arise frequently, but when they do, they correlate; moreover, the strength of that correlation is a function of the severity of the disease. As a result, epistatic, rather than additive, effects tend to provide the most likely explanation for sufficiently severe disease. The level of severity (fitness loss) required for this is sufficiently small that causal alleles may even rise to relatively high frequencies (e.g. $>1\%$).

As we found when exploring the dynamics of populations as a whole, the degree of phenotypic noise and the interaction parameter, k, have important effects on the degree to which epistasis and disease correlate. Interestingly, when k is small, it has a greater effect on the prevalence and magnitude of epistasis within a population (S6 Fig) than it does on localizing the population to the corner (S4 Fig). This effect arises because the local shape of the small portion of the landscape's corner actually occupied by the population, which depends sensitively on k, matters more for the epistasis realized in individuals of the population than for the population dynamics as a whole. Put another way, moderate nonlinearities are sufficient to localize the population to the most nonlinear part of the landscape, but large nonlinearities are needed to generate strong epistasis in severely affected cases.

Accounting for the Most Severe Common Diseases

We have seen how strong epistasis can arise in severely affected cases (those with low fitness), but how does the disease severity that arises in our simulations compare to that measured in common, heritable diseases? Averaging fitness (and epistasis) over all cases at each sampled time point reveals that diseased individuals in our simulations are most likely to have an average fitness of at least 0.8, relative to that of healthy controls (Fig 6H). For many clinically significant disorders, fitness is very likely in this range (e.g. bipolar disorder in females), but some common diseases of particular interest to human geneticists affect fitness much more severely, e.g. the average relative fitness of individuals with schizophrenia is 0.23 for men and 0.47 for women [27].

One candidate explanation for such severe, yet common, disease is rooted in the hypothesis that rare gene variants cause complex disease [28–30]. According to this hypothesis, there is a large, heterogeneous class of susceptibility variants at genes associated with disease, effectively elevating the overall mutation rate (e.g. above what we typically use in simulations), thus increasing the proportion of unfit individuals at mutation-selection balance. While S8 Fig confirms the predicted effect of increased mutation rate on the proportion of unfit individuals, it is accompanied by a sudden and drastic reduction of epistasis, presumably associated with a destabilization of the corner (cf. Fig 2D, which exhibits a similar corner destabilization when mutation step size is increased). Explaining the most severe common diseases by invoking a high effective mutation rate, it would seem, largely rules out epistasis as an important factor in the molecular underpinnings of those diseases. Accordingly, it is worth considering whether there are any other ways to explain the occurrence of such diseases at high prevalence in a population, that do not make the existence of strong epistasis unlikely.

One possibility relates to the potentially highly polygenic nature of some disease phenotypes, i.e. the degree to which genetic variants at a large number of different genes are collectively involved in causing disease. So far we have considered only an LP landscape defined by just two genes, in which only pair-wise epistasis can evolve. Yet, there is by now considerable evidence in favor of the existence of higher-order epistasis both in experimental systems [11,31–33] and humans [34]. Because the average fitness disadvantage associated with any individual allele would be expected to go down as the number of alleles with which it must interact goes up, one might expect that deleterious, high-order epistatic interactions would be less likely to be eliminated by natural selection than low-order ones, with the result that more severe epistatic disease could persist.

To test this idea, we evolved populations on an LP landscape defined by an arbitrary number of input values, each of which was determined by a different genetic locus. S9 and S10 Figs show that increasing the number of inputs (and genes) beyond the two considered thus far does not significantly alter relative case fitness.

Not only do evolutionary simulations subject to a larger number of causative genes show no increase in disease severity, they also demonstrate that synergistic epistasis between three or more genes doesn't arise spontaneously, despite the fact that the phenotypic landscapes being considered allow for epistasis among as many as six genes at a time. This surprising result is most apparent as we vary levels of total phenotypic noise, as in S9 Fig. These data show that, while it is possible to identify conditions in which case-control phenotype differences are almost entirely due to epistasis, i.e. $<H_{ind}-h_{ind}> > 0.8$, and it is possible to localize the population to the corner with respect to all input values, it is not possible to achieve both simultaneously. Instead, the intermediate levels of phenotypic noise that favor epistasis (S9A Fig; see also Fig 6E) are insufficient to corner-localize the population with respect to more than 2–3 genes at a time (S9B Fig). In summary, at mutation-selection balance, increasing the number of inputs (i.e. the number of potential interacting genes) contributing to the LP model does not increase the

contribution of epistasis to explaining disease phenotypes. However, if there were other forces keeping the population localized to a higher-dimensional corner even at modest noise levels, increasing the number of interacting genes could lead to an increase in disease epistasis.

We next turned to considering out-of-equilibrium scenarios. It has been suggested that severe disease often originate from a recent environmental shift, to which modern populations are still adapting [35]. Since our simulations already model environmental fluctuations as coordinated changes to input values, one way to simulate an abrupt environmental shift is to compute case relative fitness immediately following a large perturbation to the input values. S11 Fig shows the distribution of average case fitness (and epistasis) under such a scenario for a variety of shift magnitudes. As expected, these perturbations tend to reduce case fitness, but the effect saturates at large shift magnitudes because both cases *and* controls suffer fitness reductions. We reasoned that this saturation of disease severity would be circumvented if, instead of input value, we shifted selection strength, *s*, upward. The consequent narrowing of the ridge in the fitness landscape should leave optimal-fitness controls unscathed while dramatically reducing case fitness. As expected, a 5-fold increase in selection strength transiently increases disease severity (compare Fig 6I with 6H), attaining values comparable to those seen in schizophrenia at larger-fold selection shifts (S12 Fig). Importantly, performing the environmental shift at the level of fitness, instead of phenotype, implies that our prior conclusion—that epistasis is enriched in the most severely affected cases—remains intact (Fig 6I).

Beyond LP Models

The LP model was introduced as a mathematical representation of biochemical scenarios in which the quantity of interest is the concentration of a complex consisting of two (or more) gene products [23]. It is, however, mathematically equivalent to any scenario in which the lesser of two (or more) quantities determines function, as may easily occur when cellular components, networks, or even cells have to work together. Moreover, fitness landscapes similar to those produced by the LP model result from a variety of more general cases of gene interaction, particularly when there are tradeoffs between what genes accomplish collectively versus separately. For example, if we have two interacting gene products, X and Y, with fitness being increased by the amount of the product of $[X]$ and $[Y]$, but then also decreased separately by $[X]$ and $[Y]$ (for example, if there is some energetic cost associated with expressing both X and Y), then we easily produce landscapes such as that in Fig 7A ($Z = XY/(X+Y)$), which greatly resembles those in Fig 3A and 3B. Yet another way to produce a similar landscape is to have two interacting processes each of which is the outcome of a process that is a saturable function of a single gene (i.e. $Z = (X/(1+X))^*(Y/(1+Y))$).

Such landscapes form continuous ridges in fitness-space because they represent scenarios in which fitness tradeoffs create a continuum of options for achieving a single objective. Because tradeoffs are thought to be such an important force in biological evolution, we think it particularly relevant to understand how populations evolve on ridge-landscapes. To approach this problem in the most general way, we began by finding criteria that could describe the local shapes of all possible ridge-landscapes in three dimensions (i.e. when fitness is a function of two genes, X and Y). We identified three such criteria, which we refer to as orientation, curvature, and fall-off. The first two are characteristics of the X-Y curves of constant fitness value, i.e. the fitness contours, with orientation meaning the slope (in the X-Y plane), and curvature being the inverse of the radius of curvature. Fall-off refers to how fitness changes as one moves orthogonal to a fitness contour (i.e. the spacing between contours). It may be characterized by the degree to which it is or is not locally uniform (spacing between contours relatively constant) or locally symmetric (fitness declining equally on both sides of the optimal fitness contour).
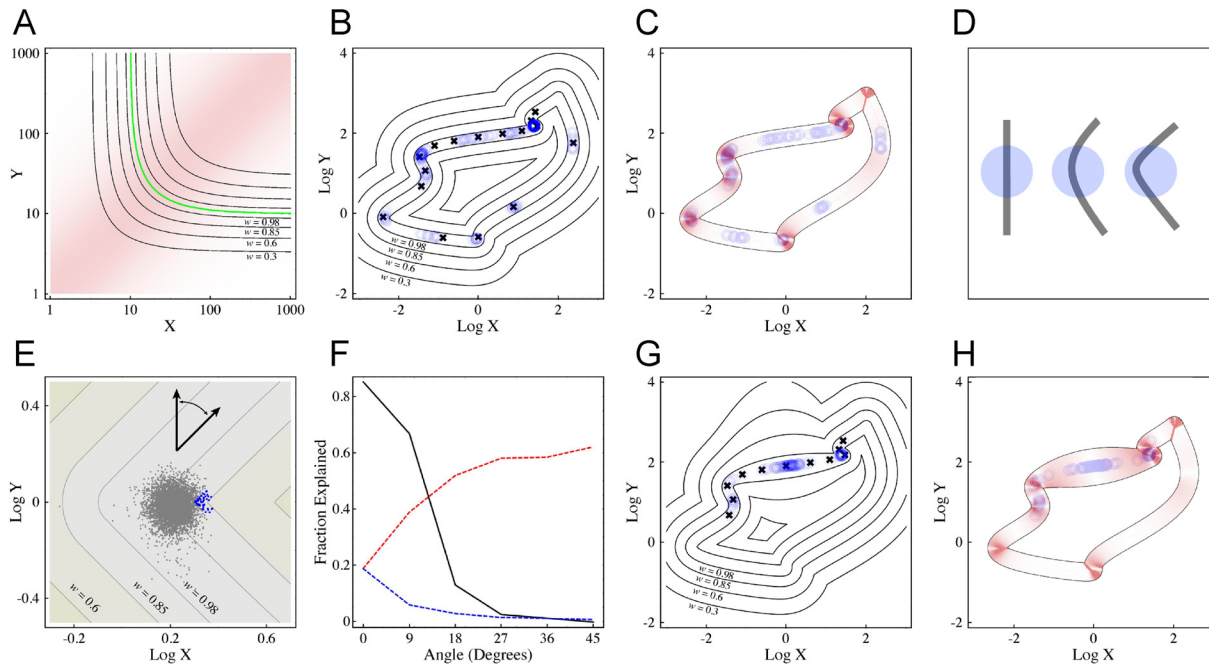
Fig 7. Evolutionary trajectories and epistasis on fitness ridges. (A) A variety of biologically realistic scenarios produce landscapes with continuous fitness ridges that resemble those in the LP model. Here we show a landscape for $Z = XY/(X+Y)$, which models a scenario in which two gene products work together, but with a cost associated with expressing each. Compare with Fig 3A and 3B. (B) Evolution on an arbitrary fitness ridge-landscape, in which the optimal phenotype ($w = 1$) is defined by a parametric curve that traces a closed loop in the X-Y plane: $\ln(X) = \sin(2t)\cos(2t) - 2\cos(t)$, $\ln(Y) = (\cosh(t) + 5\sin(t))/4$, $-\pi < t < \pi$, and for points elsewhere in the plane, fitness falls off as a function of the shortest distance to the parametric curve. Black crosses indicate initial positions of populations chosen for 22 independent simulations. A combination of developmental and environmental noise was used ($\sigma_{dev} = \sigma_{env} = 0.05$, as in Fig 6C). Blue circles show, for each simulation run, the population mean values of $\ln(X)$ and $\ln(Y)$ at 11 timepoints, sampled once every 500,000 generations between 5,000,000 and 10,000,000 generations. The results show that corners in the fitness ridge function are attractors: populations that start at a corner tend to stay there, while those starting near enough to a corner are drawn there. At three locations along the curve, there are two corners in close proximity: $(\ln(X), \ln(Y)) \approx (-1.5, 1.25)$, $(0, -0.75)$, and $(1.5, 2)$. Nearby populations are always drawn more strongly to the sharper of the two corners. (C) Maximum epistasis in the vicinity of a high-fitness ridge. This figure shows only the region of the fitness landscape from Fig 7B where $w > 0.95$. Locations of blue circles are identical to Fig 7B, representing sampled population locations at late time points. The red shading indicates the maximum possible epistasis at any given point, derived from Eq. (S85) in S1 Text. Epistasis is maximized in the vicinity of corners in the fitness ridge, and minimized where the X-Y slope (orientation) of the ridge is close to vertical or horizontal. (D) Conceptual explanation for the attractiveness of corners. Because noise forces populations to sample from nearby regions of the fitness landscape, its effect may be viewed as a form of local averaging. As sufficiently thin fitness ridges may be conceptualized as segments of curves in the X-Y plane (black), the average value of fitness within any region (blue) may be equated with the arc-length of the optimal fitness contour that is contained within it. For simple regions, the greater the curvature of the contour, the more arc length can usually be contained within it. (E, F) Influence of ridge orientation on epistasis. In D, an LP-like fitness landscape ($s = 1$; $z_{opt} = 1$, as in Fig 6C) was rotated by 45°, so that its arms become diagonal, while the orientation of the corner (slope of the optimal fitness contour at the corner) becomes vertical (the angle of rotation is indicated by the arrows). A snapshot of an evolutionary simulation ($\sigma_{dev} = \sigma_{env} = 0.05$) is shown; blue and gray dots representing "case" and "non-case" individuals, respectively. As expected, the population localizes near the corner, but the individuals with the highest phenotype values (cases) are outliers only in X relative to the rest of the population. Panel E summarizes a variety of simulation runs in which the same fitness landscape was gradually rotated from its original orientation (0°) to 45°. The black curve gives the fraction of case-control phenotype differences attributable to epistasis (for cases with relative fitness < 0.9; i.e. the same quantity plotted in Fig 6E and 6G). The red and blue dashed lines indicate the mean fraction of the phenotypic variance of the population that is attributable to variance in X and Y, respectively. Note that epistatic explanation for disease vanishes as the corner orientation moves from diagonal to vertical. At the same time, phenotypic variance goes from being bi-genic to effectively monogenic. (G) Effect of variable fitness fall-off. In this simulation, the fitness landscape in B was distorted by spreading apart the fitness contours in a relatively straight part of the landscape lying between two corners. The meanings of black and blue symbols are as in panel B. The results show that locally flattening the fitness landscape destabilizes the corner at (-1.5, 1.25). However, the sharper corner near (1.5, 2) continues to be a strong attractor. The evolutionary simulations presented in this figure are described in greater detail in Methods. (H) Maximum epistasis in the vicinity of a high-fitness ridge. The figure presents a map of maximum possible epistasis as a function of location along the high fitness ridge in panel G, using the same approach as in panel C. Here we see that the locations to which populations are drawn by "survival of the flattest" may display little or no potential for epistasis.

doi:10.1371/journal.pgen.1006003.g007

We began by investigating arbitrary landscapes in which fall-off is symmetric and relatively uniform along the entire optimal fitness contour (the LP model falls into this category), and later explored the effects of varying fall-off. Populations were initiated at various locations, and evolution simulated in the same way as was done for the LP model (see Methods). Parameters

that were varied included the strength of developmental and environmental noise, the strength of selection, and the mutational step size.

Fig 7B shows the results on one such landscape, in which the fitness contours are displayed. Black crosses show the X-Y values at which populations were initiated, while blue circles show population mean locations 5,000,000 generations later, and for a series of 500,000-generation time points thereafter. From a large number of analyses such as these we noticed two consistent trends.

First, in the presence of noise, evolutionary stability is closely related to the curvature of the optimal fitness contour, with populations typically being drawn to locations of highest curvature. These locations correlate strongly with locations at which the potential for epistasis is greatest (Fig 7C). This is the same behavior we observed in the LP model, in which noise drives populations to the corner, and, generally speaking, the sharper the corner (which corresponds to lower values of the parameter k in the LP model), the stronger the effect. We may rationalize this behavior by the fact that phenotypic noise enforces a form of spatial averaging over a region of the fitness landscape (as described above and in S1 Text, section 3). On an arbitrary fitness ridge of uniform width, "effective fitness" will depend on the area under the ridge that is contained within a given averaging radius, which will naturally be a function of the ridge's curvature. In general, the greater the curvature, the more arc-length of the optimal fitness contour will be contained within the averaged area (Fig 7D). Thus, the notion that evolution drives populations to curves—i.e. "survival of the curviest"—can be expected to apply broadly, not just to LP models or landscapes that resemble them.

Second, we also found that, as in the LP model, the emergence of epistasis as a major factor in determining the phenotypes of low-fitness individuals correlated with the arrival of populations at regions of high curvature. But, this correlation was no longer perfect. At some corners, even sharp ones, little epistasis was observed. A deciding factor, we realized, was the orientation (slope) of the optimal fitness contour (at the point of highest curvature). When populations were at locations where orientation was approximately diagonal (as it always is at the corner of LP landscapes), a large proportion of the phenotype of the most unfit individuals could be explained by epistasis. When orientation was close to horizontal or vertical, almost none was. We can rationalize this behavior by noticing that the slope of the optimum fitness contour is a measure of the relative contributions of changes in X and Y to phenotype. When that slope parallels one of the two axes, phenotype is necessarily insensitive to small changes along that axis. Epistasis between X and Y becomes negligible because phenotypic variation correlates significantly with genetic variation at only one of the two loci. We can illustrate this graphically (Fig 7E and 7F) by rotating the LP landscape about the origin to various extents, evolving populations on them, and observing that, as the fraction of the phenotypic variance explained, individually, by X and Y goes up toward one, the fraction of the case-control phenotype differences attributable to epistasis, for cases with relative fitness less than 0.9 (the same quantity plotted in Fig 6E and 6G), goes down toward zero. As described in S1 Text, we can more generally analyze the effects of orientation and curvature on epistasis by deriving general expressions for orientation, curvature, and maximum possible epistasis in terms of the partial derivatives of the phenotype with respect to X and Y. Such analysis reveals the generality of the relationship between orientation and potential for epistasis, and specifies the exact conditions under which curvature and epistasis will correlate with one another.

After verifying these findings over a variety of arbitrary ridge shapes, we considered the effects of varying fall-off. As shown in Fig 7G, by appropriately adjusting the spacing between fitness contour lines, it is possible to find scenarios in which evolution in the presence of noise drives populations away from regions of highest curvature, and toward regions of low potential for epistasis (Fig 7H). Such simulations reveal that evolutionary dynamics in the presence of

noise reflects two "forces" that can be at odds with each other: survival of the curviest, as described above, and "survival of the flattest". The latter has been described previously, as a means of explaining how evolution selects for robust, or canalized, outcomes [36]. Like survival of the curviest, survival of the flattest arises from the fact that, in the presence of noise, "effective fitness" reflects a local averaging of the fitness landscape. A population may increase that average fitness either by moving to a landscape location that is curvier (Fig 7D) or flatter (Fig 7G), depending upon which feature exerts the stronger attraction, but only in the former case may such movement be expected to produce a strong increase in epistasis. Thus, even though the conclusions drawn here from the analysis of LP landscapes may apply to a broad range of biological scenarios, we should not expect them to apply to all (see Discussion).

Discussion

Human populations are ill-suited for the direct, empirical discovery of epistasis [6], which is why theory and simulation currently play a major role in the debate about whether epistasis is an important factor in the genetic etiology of human variation, e.g. [37,38]. Here we used a combination of mathematical analysis and evolutionary simulation to investigate the conditions under which the evolutionary forces of drift and natural selection tend to cause relatively common, deleterious traits ("common disease") to be due substantially to epistasis. A central aspect of our analysis was the inclusion of phenotypic "noise", i.e. randomness in the genotype-phenotype map either at the level of individuals (developmental noise) or at the level of the map itself (environmental noise).

It has long been appreciated that noise can favor the evolution of canalization [39,40], although the relationship between, and relative importance of, different kinds of noise continue to be debated [41–43]. Moreover, noise-driven canalization has previously been observed to correlate with an increase in epistasis. For example, when evolving digital organisms *in silico*, Wilke and Adami noted that at sufficiently high mutation rates—effectively a form of long-term phenotypic noise—populations tended to be drawn to flatter regions of the phenotypic landscape ("survival of the flattest"), even if such regions were suboptimally fit, and this was accompanied by an increase in the prevalence of synergistic epistasis [44]. Qualitatively similar findings of mutational robustness emerging together with epistasis have come from simulated evolution of gene regulatory networks in the presence of recombination [15], and have been confirmed in experimental studies of the function of single enzymes in bacteria [45].

Because noise in the genotype-phenotype relationship may be viewed as a form of local averaging of the fitness landscape (see S1 Text), it is not surprising that several different kinds of noise—including developmental and environmental variability of the kind we consider here—have all been observed to drive evolving populations to flatter locations [46–48], where robustness to variation of any sort (genetic or environmental) will be buffered [47,49]. What has been unclear is whether such movement is necessarily accompanied by an increase in epistasis, and if it is, whether an identifiable relationship exists between the parameters of noise, the phenotypic landscape, and distribution of epistatic effects that arise among individuals.

In part to address these questions, we studied a class of simple landscapes defined by the generalized LP model, which may be viewed as an abstraction of several biologically relevant, mathematically related, gene-gene interaction scenarios. Rather than consider only average epistatic effects, we focused on the necessary conditions for especially unfit ("disease") phenotypes to be explained by epistasis. These conditions included intermediate levels of phenotypic noise (Fig 6E), selection strength (Fig 6G), and mutation rate (S8 Fig), as well as a sufficiently nonlinear genotype-phenotype mapping (i.e. small k; S4 Fig). The extent to which selection drove populations to highly epistatic regions of the phenotypic landscape depended on the

magnitude of the non-linearity (e.g., parameter k in S4 Fig), as well as the scale over which the population was spread—which depended on both the strength of selection, s, and phenotypic noise level, σ (through the product $s\sigma^2$; Fig 6G).

When these conditions were met, synergistic interactions among both rare and common alleles (S5 Fig) tended to be a major cause of disease, particularly in the most severely affected individuals (Fig 6C), regardless of the number of genes underlying the trait in question (S10 Fig). That the most severely affected individuals are expected to be the most "epistatic" can be understood by noting that such individuals are farthest away, on the phenotypic landscape, from the norm, and therefore most influenced by the landscape's nonlinearity. In agreement with this prediction, we note that, in recent yeast work that has identified substantial amounts of epistasis among standing genetic variants, analysis had been focused specifically on the most extreme phenotypes [11]. Although the individuals that possess such phenotypes may be rare, they can be better positioned to provide mechanistic insights into the phenotypic landscape on which an entire population resides.

When we subsequently extended our analysis to other kinds of two-gene fitness surfaces, all of which, like the LP model, had optimal-fitness ridges, we again observed evolving populations drawn by noise to the most sharply curved parts of ridges, a phenomenon we term "survival of the curviest". Arrival at highly curved regions tended to correlate with an increase in robustness, as well as reduced phenotypic variance despite increased accumulation of genetic variation. Furthermore, as in the LP model, arrival at such locations also tended to correlate with an increase in observed epistasis. This was not always the case, however, as this pattern was violated in two types of situations:

First, when ridge corners (vertices) were oriented close to horizontal or vertical, populations were still drawn to them, but strong epistasis was not observed, as phenotypic variance essentially became a function of only one of the two genes (Fig 7E and 7F). In practical terms, localization of a population to this type of landscape feature ought to make the discovery of causal gene variants by traditional statistical means (e.g. GWAS) much easier, because it both reduces the number of genes at which variation has any effect, and eliminates all but "main effects". Accordingly, to the extent that one cares primarily about heritable diseases that GWAS has failed to adequately explain, one should be biased against encountering this type of landscape feature.

Second, when we varied the way fitness declined along ridges, we found that "survival of the curviest" could sometimes be thwarted by "survival of the flattest", with populations being drawn to locations at which two genes still contribute to phenotypic variance, but little epistasis occurs (Fig 7G and 7H). Such simulations show that the connection that has previously been noticed between selection for robustness and epistasis (e.g. [15,44]) is not, in fact, a necessary one (see also S1 Text, section 6, for an analytical treatment), and that our ability to generalize conclusions drawn from the study of LP landscapes to other biological settings depends upon whether landscapes in which flatness outweighs curviness are found frequently or rarely in biology. We note, in this regard, that the landscapes explored in Fig 7 were arbitrary curves, not meant to be representative of any specific biological processes. How easily networks of actual biological components achieve such topographies remains to be investigated.

It should be noted that none of the conclusions of the present study are necessarily in conflict with earlier work on the evolution of epistasis [15–18], including studies that have suggested that epistasis is unlikely to play a significant role in complex human traits in general, e.g. [14]. The reason is that we have focused here not on traits in general, but on a very particular subset, namely those that are substantially deleterious (e.g. fitness <0.9) and moderately (but not very) common (population frequency between 0.1 and 10%). We justify this focus by the fact that such criteria would appear to capture a great deal of clinically relevant human disease.

There are two reasons why this focus is so important. First, the situations in which both criteria (substantially deleterious and moderately common) are met are evolutionarily rare events (because natural selection disfavors the appearance of common disease). Accordingly, arguments based on the typical evolutionary behaviors of traits do not necessarily apply. In effect, we are not asking here about the probability of observing epistasis in general, but on its probability given prior knowledge of common disease.

Second, by restricting our view to disease frequencies of 10% or lower, our analysis operates in a regime in which functional epistasis has the potential to exist without much statistical epistasis (Fig 1). This means that the effects of epistasis on the evolutionary process itself can be very small (deflating the argument that epistasis *per se* is deleterious, and thus should be selected against, e.g. [37]). It also means that the impact of epistasis on patterns of heritability can be very small (which could likewise eliminate most arguments against the existence of epistasis based on the analysis of data, e.g. GWAS). The condition under which functional epistasis can be large, while statistical epistasis is negligible, is when populations are distributed on a phenotypic landscape that is relatively linear where the majority, "control" individuals are found, becoming non-linear where the minority, "cases" reside (see Fig 1). On the simple LP landscape, such a scenario occurs when populations are centered near the "corner", i.e. when inputs from different loci are reasonably balanced. As we show here, the force that drives populations to that location is noise.

In the realm of human genetics, the use of unbiased genetic association studies to uncover explanations for complex human disease has often proved exceptionally challenging, leaving clinicians and researchers to decide how best to allocate further efforts. Prominent competing hypotheses have attributed undiscovered heritable variation variously to epistatic interactions among genes; to large numbers of variants of very small effect; or to variants currently excluded from the association studies. The latter include both rare variants and structural alterations other than SNPs, such as tandem-repeat polymorphisms [50], copy-number variants [51], and large translocations [52]. While such structural variants have been associated with individual diseases (e.g., [53]), they have yet to account for a great deal of the "missing" heritability of common diseases.

Here, by examining the conditions that favor the evolution of disease epistasis, we hoped to illuminate circumstances under which epistatic explanations are most likely. We believe that those circumstances, while not universal, are likely common enough to warrant additional effort toward discovering epistatic interactions. In this regard, we can imagine at least three specific ways in which the present study might be useful. First, it may help investigators use information about the social, cultural or historical aspects of a particular disease state to develop intuitions about the probability of an epistatic explanation. For example, expected interactions between mental illness and culture, or metabolic illness and diet, suggest that diseases in these categories are more likely to be evolving in the face of recently fluctuating environments than, say, congenital heart defects. Second, it could encourage investigators to directly measure developmental noise genome-wide, e.g., as reflected in variations in RNA-seq levels in tissues from individuals with the same SNP genotype [54], with the goal of identifying cases in which the observed noise is of the right magnitude to favor the emergence of epistatic, disease-causing interactions. Third, our simulations of the LP model suggest that risk alleles that combine to produce exceedingly unfit individuals (fitness < 0.75) ought to have frequencies that lie in the 0.2–0.5% range, at the low end of what is typically considered in GWAS studies, but higher than expected for rare alleles with Mendelian effects. Potentially, GWAS that focused exclusively in this allele frequency range might actually have greater power to detect epistasis than GWAS that considers all variants.

Methods

Numerical calculations of linear regression were implemented in python and visualized in Matlab. Evolutionary simulations were implemented in C, as described below.

We simulated populations of $N = 5000$ diploid individuals forward in time using non-overlapping generations. An individual's genotype consisted of two allele values at each locus. Each input value (e.g., x or y) was determined by a single locus, and was taken as the sum of the individual's two allele values at that locus. We restricted allele values to be less than 100 (i.e. each input value was constrained to be less than 200). Every generation, then, each individual's allele values were summed to determine its nominal input values. These values were adjusted first by environmental noise, which exerts the same effect on each individual in the population. The resulting values were then adjusted by developmental noise, applied independently to each individual.

The following Metropolis-Hastings algorithm was used to model persistence of environmental noise from one generation to the next, i.e. the temporal autocorrelation of the noise. Given that the environmental perturbation term in generation i was Δx_i, the perturbation in the next generation, Δx_{i+1}, was determined as follows. A proposed value of the perturbation, Δx_{prop}, was generated by adding a Gaussian-distributed random variate with mean zero and variance σ_{step}^2 to Δx_i. If $\Pr(\Delta x_{prop}) = exp[-\Delta x_{prop}^2/(2\sigma_{env}^2)]$ was greater than $\Pr(\Delta x_i) = exp[-\Delta x_i^2/(2\sigma_{env}^2)]$, then the proposed perturbation was accepted and Δx_{i+1} was assigned the value Δx_{prop}. Otherwise, the proposed perturbation was accepted with probability proportional to the ratio $\Pr(\Delta x_{prop})/\Pr(\Delta x_i)$. If the proposed perturbation was rejected, then the environmental perturbation remained unchanged (Δx_{i+1} was assigned the value Δx_i). This algorithm ensures that the steady-state probability that the environmental perturbation has a particular value, Δx, is proportional to $exp[-\Delta x^2/(2\sigma_{env}^2)]$. For given values of σ_{step} and σ_{env}, the autocorrelation time was determined *a posteriori* from the autocorrelation function of Δx_i. For most of the results presented here, $\sigma_{step}/\sigma_{env} = 0.25$, generating an autocorrelation time of ≈ 27 generations.

Having applied the phenotypic noise to the nominal input values, the resulting set of effective input values (x_e, y_e) was used to calculate the individual's phenotype, z, and fitness, w. Simulations with $d > 2$ genes were performed in an analogous way, with the effective set of input values ($x_{1e}, x_{2e}, \ldots x_{de}$) determined by the application of environmental and developmental noise to each of the d nominal input values ($x_1, x_2, \ldots x_d$).

The next generation was then constructed by a method of sampling with replacement, where the probability of an individual being sampled to become a parent was proportional to its fitness. For each new offspring, two parents were identified using the following sampling-rejection scheme. First, a prospective parent was drawn at random from the entire population. Then, a uniform random number was drawn between zero and the highest fitness in the population. If the prospective parent's fitness was higher than the random number, the parent was accepted. If not, they were rejected and a new random prospective parent was drawn. Sampling with replacement implies that a rejected parent could be redrawn and accepted later, and a parent could be drawn more than once; the only exception was that the same candidate could not be drawn for both parents of the same individual.

The offspring genotype was constructed by randomly selecting one of the two alleles at the first locus from each parent. A recombination parameter determined the probability of switching chromosomes between loci. In most simulation results presented in the text, we assumed free recombination, so that the two alleles at each locus were sampled with equal probability and with no correlation among loci.

Finally, a mutation was applied independently to each allele with probability $\mu = 10^{-4}$ per allele per generation. In the event of a mutation, a normally distributed quantity (mean zero,

variance σ_μ^2) was added to the natural logarithm of the allele's value (thereby ensuring that each allele value remains positive). If the mutation led to an allele value in excess of 100, the mutation was rejected, and the inherited allele retained its original value.

Each simulation was run for 10 million generations, and the first 1 million generations discarded. Populations were typically analyzed at intervals of 10,000 generations.

At the start of each simulation on the generalized LP landscape, each individual was homozygous, with allele values chosen so that $x = y$ and $z = 1$, i.e. all individuals were positioned at the corner of the ridge in the phenotypic-noise-free fitness landscape. For example, when $k = 0$, the initial value associated with each allele was 0.5, so that $x = y = 1$.

For evolutionary simulations on arbitrary, non-LP landscapes (Fig 7), a lookup table of phenotype values was pre-computed based on a grid of combinations of log-transformed underlying gene values ($\ln(x)$ and $\ln(y)$). During the course of simulation, individual phenotype values were calculated by linear interpolation from the four closest $\ln(x)$, $\ln(y)$ combinations from the lookup table. Fitness values were then calculated according to (Eq 2) in the main text.

For Fig 7B and 7C, the phenotype lookup table was constructed as follows. An array of pairs, ($ln(x)$, $ln(y)$) was constructed consisting of 2100 values of $\ln(x)$ from -3.5 to 3.5 by 1800 values of $ln(y)$ from -2 to 4. For each pair, the logarithm of the phenotype value z was taken to be a multiple of the distance d to the closest point on the parametric curve corresponding to the optimal phenotype, $z_{opt} = 1$. For points outside the closed curve, $ln(z) = -d$, so that $z < z_{opt}$; for points inside the closed curve, $\ln(z) = d$ so that $z > z_{opt}$.

For Fig 7G and 7H, the phenotype lookup table was constructed as above, but with the following modification. When calculating the distance d, we introduce a scaling factor f that depends on the curve parameter t (defined by the point on the curve nearest to ($\ln(x)$, $\ln(y)$)). For $t < \pi/3$ or $t > 2\pi/3$, f is set to 1. But for $\pi/3 < t < 2\pi/3$, f is set to $(3 - \cos(6t))/2$. This reduces the calculated distances in the vicinity of $t = \pi/2$, meaning that phenotype and fitness values change less rapidly in response to changes in x and y.

Supporting Information

S1 Fig. Measures of epistasis. (A, B) Low statistical epistasis can exist in the presence of high functional epistasis. Statistical epistasis was computed using either all of the genotypes in a population (red), or a biased version of that distribution in which the prevalence of individuals with large phenotype values was deliberately increased (green). In this example, the prevalence of individuals displaying the top 1% of phenotypes was increased to 50%, as would typify a balanced case-control study of a common disease. Eq (S34) defines the mathematical relationship between the distributions of genotypes in the original and adjusted populations. Panels A and B differ in the functional form of the phenotypic landscape, corresponding to different values of the parameter n in the definition, $Z(x, y) = x + y + \frac{(xy)^n}{\xi^{2n-1}}$; the lines in each panel were generated by varying the parameter ξ, which controls the allele values at which the landscape crosses over from relatively linear to highly nonlinear. The population distribution was defined by $p(x, y) = \frac{1}{\lambda^2} e^{-(x+y)/\lambda}$ with $\lambda = 0.2$. Functional epistasis was averaged over all case individuals. The results show that enriching for case genotypes, which have the greatest levels of functional epistasis, does not necessarily increase statistical epistasis, and often reduces it. (C) Though functional epistasis of cases is not sufficient to generate high levels of statistical epistasis (panels A and B), the latter can result from moderate levels of functional epistasis in the bulk of the population, as measured by evaluating functional epistasis at one standard deviation of the (original) genotype distribution, $1 - h_{ind^*}(\lambda, \lambda \mid 0,0)$. The data also show that statistical epistasis depends on the spread of the population, λ, and the shape of the phenotype landscape, ξ, only through the lumped parameter, $1 - h_{ind^*}(\lambda, \lambda \mid 0,0)$ (derived in

Section 1.2 in SI text). Line is defined by Eqs (S27)–(S29); circles are numerical estimates for a set of (ξ, λ) values. $n = 2$.
(PDF)

S2 Fig. Distribution of nominal and effective input values under developmental and environmental noise and under case-control sampling. (A-D) Distribution of input values among all individuals (green) in a simulated population and conditioned upon individuals being cases (pink) or controls (dark blue). Controls are drawn from the middle 50% of rank-ordered phenotype value (ranging from the 25th to the 75th percentile); cases represent the top 1% of this rank-ordered distribution. Since developmental noise is uncorrelated among individuals, the *nominal* input values of cases and controls are poorly differentiated (A), implying little heritability. Environmental noise, on the other hand, is correlated across individuals, resulting in a correlation between *nominal* input values and case-control status (C), and high heritability. (E-H) Time courses. Blue lines connect points corresponding to the average values of x and y in the population at intervals of 500,000 generations. Red bars indicate the magnitude of the variation in the population at each sampled timepoint (+/-2 standard deviations). (I-L) Median locations of case (pink) and control (dark blue) samples for the same temporal sequence of samples shown in panels E through H. Black lines connect control and case medians from the same time point. Parameter values are $z_{opt} = s = 1$, $N = 5000$, $k = 0.01$, and $\sigma_\mu = \sigma_{dev} = \sigma_{env} = 0.1$
(PDF)

S3 Fig. Random walk on the LP landscape. (A) Evolutionary path of the population center of mass as a function of input values. Selection is governed by the LP fitness landscape, with hotter colors (white, yellow) and colder colors (red, black) representing high and low fitness values, respectively. When interaction strength is weak ($k>>1$), as shown here, the corner is evolutionarily unstable. (B) When the interaction strength is strong ($k<<1$), the population follows an unbiased random walk along the arms, except in close proximity to the corner, where it tends to localize. (C) The random walk of the population center of mass along the horizontal arm is biased towards the corner by the graded effective fitness profile, $w_e(x)$, where x represents the natural logarithm of the input value. (D) $g(x|x')$ is the probability density that a mutation changes (the natural log of) the value of the gene to x, given that (the natural log of) the input value prior to the mutation was x'. (E) The probability, $\rho(x|x')$, that a mutant allele with value x fixes in a population of alleles of value x' (again in log space), corresponding to the effective fitness profile shown in (C). In C–E, the horizontal axis represents the *natural logarithms* of the corresponding input values. N is population size.
(PDF)

S4 Fig. Phenotypic noise stabilizes the corner at smaller k. Fraction of simulations in which the population-mean ratio of the two input values lay within the indicated bounds (1.2 for A and B, 1.4 for C and D, and 3.0 for E and F). Parameter values were $z_{opt} = s = 1$ and $N = 5000$. At moderate noise levels (A, C, and E), environmental and developmental noise behave similarly in terms of localizing the population at the corner. At higher noise levels (B, D, and F), developmental noise localizes the population to the corner more effectively than does environmental noise.
(PDF)

S5 Fig. Population frequencies of alleles causing disease. Each case individual is represented as a single point whose color indicates the population frequency of the rarest allele carried by that individual. (Recall that the phenotype value z depends on the values of an individual's two alleles at each of two loci.) Results are pooled over the five timepoints in one simulation run

with the lowest mean fitness of cases relative to controls. Parameter values are $\sigma_{env} = \sigma_{dev} = 0.05$, $s = 1$, and $k = 0$.
(PDF)

S6 Fig. How epistasis depends on the interaction constant, *k*. Heat maps with intensity representing the epistatic fraction, H_{ind}—h_{ind}, averaged over all cases with relative fitness < 0.9 and all sampled time points for $k = 0$ (A) and $k = 0.01$ (B). The combinations of environmental and developmental noise levels that lead to significant epistasis are similar in the two cases, but stronger interaction (i.e. lower value of k) results in a greater fraction of the case-control phenotype difference being attributable to epistasis.
(PDF)

S7 Fig. Low relative fitness predicts epistasis among cases. Summary of the data in Fig 6C. Circles indicate the median value of epistasis fraction, H_{ind}—h_{ind}, for case individuals with similar fitness values, and error bars indicate the corresponding 25th and 75th percentile values.
(PDF)

S8 Fig. How mutation rate, *μ*, affects epistasis and case fitness. Mutation rates (per allele per generation) for each row are indicated on the right-hand side. The left-hand panels show the joint distribution of the epistatic fraction, H_{ind}—h_{ind}, and fitness (relative to mean control fitness) over all cases and all sampled time points. In the right-hand panels, each dot represents a single time point at which the epistatic fraction and relative fitness were averaged over all cases. Parameter values are $\sigma_{dev} = \sigma_{env} = 0.05$ and $\sigma_{\mu} = 1.0$.
(PDF)

S9 Fig. Effect of dimensionality (number of inputs) on disease epistasis and corner localization. (A) Epistatic fraction H_{ind}–h_{ind} averaged over all cases with a relative fitness of less than 0.9 and over all time points, as a function of the total phenotypic noise level $\sigma = \sqrt{\sigma_{dev}^2 + \sigma_{env}^2}$, illustrating the biphasic dependence of epistasis on noise level evident in Fig 6E and 6G. Epistasis decreases more with increasing developmental, versus environmental, noise (compare second and third rows). Increasing the dimensionality of the system has only a modest effect on the epistatic fraction, with the largest effect occurring when $\sigma_{env} > \sigma_{dev}$. (B) The number of inputs that are localized to the corner, defined as the number of time- and population-averaged input values that are within one mutational step (σ_{μ}) of the corner, as a function of phenotypic noise level, $\sigma = \sqrt{\sigma_{dev}^2 + \sigma_{env}^2}$. Greater values of phenotypic noise are required to localize larger numbers of dimensions to the corner. As seen in Fig 5, increasing developmental noise monotonically enhances corner localization (second row) while intermediate levels of environmental noise correspond to maximum corner localization (third row). In both columns, rows differ in the relative contributions of σ_{dev} and σ_{env} to σ.
(PDF)

S10 Fig. Effect of dimensionality (number of inputs) on epistasis and case fitness. Each row corresponds to a different number of inputs in the LP model (indicated on the right-hand side). Each input value is determined by its own locus, and all loci are unlinked. The first and third columns show the joint distribution of the epistatic fraction, H_{ind}—h_{ind}, and fitness (relative to mean control fitness) for all cases and all sampled time points. In the second and fourth columns, each dot represents a single time point, at which the mean values of the epistatic fraction H_{ind}–h_{ind} and case relative fitness were calculated. The first and second column ($\sigma_{dev} = \sigma_{env} = 0.05$) correspond to the condition under which we observed the strongest epistasis among cases with relative fitness < 0.9 (the peak in the first row of S9A Fig). The third and fourth column ($\sigma_{dev} = 0.05$; $\sigma_{env} = 0.15$) correspond to the condition under which

dimensionality appeared to have the greatest influence on the epistatic fraction for cases with relative fitness < 0.9 ($\sigma = 0.158$ in third row of S9A Fig).
(PDF)

S11 Fig. Effect of sudden environmental shift on case epistasis and fitness. Simulations were run as described, but in this case an additional environmental shift was temporarily performed immediately prior to sampling case and control individuals at each time point (represented by points on the graphs). Environmental shifts were implemented by increasing both input values by a fixed quantity Δ_{env} on a logarithmic scale (*e.g.*, $\ln x$ became $\ln x + \Delta_{env}$). Shifting the environment in this way reduces the fitness of cases relative to controls, but the reduction is limited by the fact that both cases and controls suffer a fitness cost at large environmental shifts. Parameter values were $k = 0$; $s = 1$; $\sigma_{dev} = \sigma_{env} = 0.05$; $\sigma_{\mu} = 1$; $\mu = 10^{-4}$.
(PDF)

S12 Fig. Effect of a sudden increase in selection on case epistasis and fitness. Simulations were run with selection strength $s = 1$ (top row). At each time point, we temporarily increased the strength of selection from s to s' before averaging epistasis, H_{ind}—h_{ind}, and relative fitness over all cases (second through third rows). Under pure developmental noise (left column, top row), all cases have reduced fitness relative to controls, reflecting the fact that the population is centered at the corner. In contrast, under pure environmental noise (right column, top row), values of case fitness extend upwards to 1 (and beyond), reflecting the fact that the population is not necessarily centered on the corner. That is, when environmental noise shifts the population so that most of its individuals have sub-optimal phenotype values, cases, which represent the individuals with the highest phenotype values, will have higher fitness than controls. Since increasing the strength of selection has little effect on high-fitness individuals, it can have a substantial impact on the relative fitness of cases and controls (second through third rows). Parameter values were $k = 0$ and $\sigma_{\mu} = 1$.
(PDF)

S1 Text. Formulation and analysis of mathematical models.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: JFW PTM JG ADL. Performed the experiments: JFW PTM. Analyzed the data: JFW PTM JG ADL. Contributed reagents/materials/analysis tools: JFW PTM. Wrote the paper: JFW PTM JG ADL.

References

1. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747–753. doi: 10.1038/nature08494 PMID: 19812666

2. Culverhouse R, Suarez BK, Lin J, Reich T (2002) A Perspective on Epistasis: Limits of Models Displaying No Main Effect. The American Journal of Human Genetics 70: 461–471. doi: 10.1086/338759 PMID: 11791213

3. Moore JH, Williams SM (2005) Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. Bioessays 27: 637–646. doi: 10.1002/bies.20236 PMID: 15892116

4. Mackay TFC (2014) Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet 15: 22–33. Review. doi: 10.1038/nrg3627 PMID: 24296533

5. Steen KV (2012) Travelling the world of gene-gene interactions. Brief Bioinform 13: 1–19. doi: 10.1093/bib/bbr012 PMID: 21441561

6. Wei W-H, Hemani G, Haley CS (2014) Detecting epistasis in human complex traits. Nat Rev Genet 15: 722–733. doi: 10.1038/nrg3747 PMID: 25200660

7. Brem RB, Storey JD, Whittle J, Kruglyak L (2005) Genetic interactions between polymorphisms that affect gene expression in yeast. Nature 436: 701–703. doi: 10.1038/nature03865 PMID: 16079846

8. Shao H, Burrage LC, Sinasac DS, Hill AE, Ernest SR, et al. (2008) Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. Proc Natl Acad Sci U S A 105: 19910–19914. doi: 10.1073/pnas.0810388105 PMID: 19066216

9. 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. doi: 10.1073/pnas.1213423109 PMID: 22949659

10. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L (2013) Finding the sources of missing heritability in a yeast cross. Nature 494: 234–237. doi: 10.1038/nature11867 PMID: 23376951

11. Taylor MB, Ehrenreich IM (2014) Genetic interactions involving five or more genes contribute to a complex trait in yeast. PLoS Genetics 10: e1004324. doi: 10.1371/journal.pgen.1004324 PMID: 24784154

12. Hemani G, Shakhbazov K, Westra H-J, Esko T, Henders AK, et al. (2014) Detection and replication of epistasis influencing transcription in humans. Nature 508: 249–253. doi: 10.1038/nature13005 PMID: 24572353

13. Brown AA, Buil A, Vinuela A, Lappalainen T, Zheng H-F, et al. (2014) Genetic interactions affecting human gene expression identified by variance association mapping. Elife 3: e01381. PMID: 24771767

14. Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. PLoS Genetics 4: e1000008. doi: 10.1371/journal.pgen.1000008 PMID: 18454194

15. Azevedo RBR, Lohaus R, Srinivasan S, Dang KK, Burch CL (2006) Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature 440: 87–90. doi: 10.1038/nature04488 PMID: 16511495

16. Desai MM, Weissman D, Feldman MW (2007) Evolution can favor antagonistic epistasis. Genetics 177: 1001–1010. doi: 10.1534/genetics.107.075812 PMID: 17720923

17. Gros P-A, Le Nagard H, Tenaillon O (2009) The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. Genetics 182: 277–293. doi: 10.1534/genetics.108.099127 PMID: 19279327

18. Draghi JA, Plotkin JB (2013) Selection biases the prevalence and type of epistasis along adaptive trajectories. Evolution 67: 3120–3131. doi: 10.1111/evo.12192 PMID: 24151997

19. Neher RA, Shraiman BI (2011) Statistical genetics and evolution of quantitative traits. Rev Mod Phys 83: 1283–1300.

20. Keightley PD (1989) Models of quantitative variation of flux in metabolic pathways. Genetics 121: 869–876. PMID: 2721937

21. Gjuvsland AB, Hayes BJ, Omholt SW, Carlborg O (2007) Statistical Epistasis Is a Generic Feature of Gene Regulatory Networks. Genetics 175: 411–420. doi: 10.1534/genetics.106.058859 PMID: 17028346

22. Mäki-Tanila A, Hill WG (2014) Influence of Gene Interaction on Complex Trait Variation with Multilocus Models. Genetics 198: 355–367. doi: 10.1534/genetics.114.165282 PMID: 24990992

23. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences 109: 1193–1198. doi: 10.1073/pnas.1119675109

24. Karlin S, Levikson B (1974) Temporal fluctuations in selection intensities: Case of small population size. Theoretical Population Biology 6: 383–412. doi: 10.1016/0040-5809(74)90017-3

25. Bengtsson M, Ståhlberg A, Rorsman P, Kubista M (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Research 15: 1388–1392. doi: 10.1101/gr.3820805 PMID: 16204192

26. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature 459: 428–432. doi: 10.1038/nature08012 PMID: 19363473

27. Power RA, Kyaga S, Uher R, MacCabe JH, Långström N, et al. (2013) Fecundity of Patients With Schizophrenia, Autism, Bipolar Disorder, Depression, Anorexia Nervosa, or Substance Abuse vs Their Unaffected Siblings. JAMA Psychiatry 70: 22–30. doi: 10.1001/jamapsychiatry.2013.268 PMID: 23147713

28. Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69: 124–137. doi: 10.1086/321272 PMID: 11404818

29. Goldstein DB (2009) Common genetic variation and human traits. N Engl J Med 360: 1696–1698. doi: 10.1056/NEJMp0806284 PMID: 19369660

30. McClellan J, King M-C (2010) Genetic heterogeneity in human disease. Cell 141: 210–217. doi: 10.1016/j.cell.2010.03.032 PMID: 20403315

31. Pettersson M, Besnier F, Siegel PB, Carlborg O (2011) Replication and explorations of high-order epistasis using a large advanced intercross line pedigree. PLoS Genetics 7: e1002180. doi: 10.1371/journal.pgen.1002180 PMID: 21814519

32. Wang Y, Arenas CD, Stoebel DM, Cooper TF (2013) Genetic background affects epistatic interactions between two beneficial mutations. Biol Lett 9: 20120328. doi: 10.1098/rsbl.2012.0328 PMID: 22896270

33. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB (2013) Should evolutionary geneticists worry about higher-order epistasis? Curr Opin Genet Dev 23: 700–707. doi: 10.1016/j.gde.2013.10.007 PMID: 24290990

34. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, et al. (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69: 138–147. doi: 10.1086/321276 PMID: 11404819

35. Gibson G (2009) Decanalization and the origin of complex disease. Nat Rev Genet 10: 134–140. doi: 10.1038/nrg2502 PMID: 19119265

36. Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C (2001) Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412: 331–333. doi: 10.1038/35085569 PMID: 11460163

37. Kelly JK (2000) Epistasis, linkage and balancing selection. In: Wolf JB, Brodie ED III, Wade MJ, editors. Epistasis and the Evolutionary Process. New York: Oxford University Press. pp. 146–157.

38. Liberman U, Feldman MW (2005) On the evolution of epistasis I: diploids under selection. Theoretical Population Biology 67: 141–160. doi: 10.1016/j.tpb.2004.11.001 PMID: 15808333

39. Wagner A (1996) Does Evolutionary Plasticity Evolve. Evolution 50: 1008–1023.

40. Wagner GP, Booth G, Bagheri-Chaichian H (1997) A Population Genetic Theory of Canalization. Evolution 51: 329–347.

41. de Visser JAGM, Hermisson J, Wagner GP, Meyers LA, Bagheri-Chaichian H, et al. (2003) PERSPECTIVE: EVOLUTION AND DETECTION OF GENETIC ROBUSTNESS. Evolution 57: 1959–1972. doi: 10.1554/02-750R PMID: 14575319

42. Wagner A (2005) Robustness and Evolvability in Living Systems. Princeton University Press.

43. Masel J, Siegal ML (2009) Robustness: mechanisms and consequences. Trends in genetics 25: 395–403. doi: 10.1016/j.tig.2009.07.005 PMID: 19717203

44. Wilke CO, Adami C (2001) Interaction between directional epistasis and average mutational effects. Proceedings of the Royal Society B: Biological Sciences 268: 1469–1474. doi: 10.1098/rspb.2001.1690 PMID: 11454290

45. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. Nature 444: 929–932. doi: 10.1038/nature05385 PMID: 17122770

46. Kawecki TJ (2000) The evolution of genetic canalization under fluctuating selection. Evolution 54: 1–12. PMID: 10937177

47. Lehner B (2010) Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. PLoS ONE 5: e9035. doi: 10.1371/journal.pone.0009035 PMID: 20140261

48. Soyer OS, Pfeiffer T (2010) Evolution under fluctuating environments explains observed robustness in metabolic networks. PLoS Comput Biol 6. doi: 10.1371/journal.pcbi.1000907

49. Kaneko K (2011) Proportionality between Variances in Gene Expression Induced by Noise and Mutation: Consequence of Evolutionary Robustness. BMC Evol Biol 11: 27. doi: 10.1186/1471-2148-11-27 PMID: 21269459

50. Hannan AJ (2010) Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. Trends Genet 26: 59–65. doi: 10.1016/j.tig.2009.11.008 PMID: 20036436

51. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, et al. (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 464: 713–720. doi: 10.1038/nature08979 PMID: 20360734

52. Raphael BJ (2012) Chapter 6: Structural variation and medical genomics. PLoS Comput Biol 8: e1002821. doi: 10.1371/journal.pcbi.1002821 PMID: 23300412

53. Wheeler E, Huang N, Bochukova EG, Keogh JM, Lindsay S, et al. (2013) Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. Nat Genet 45: 513–517. doi: 10.1038/ng.2607 PMID: 23563609

54. Hughes DA, Kircher M, He Z, Guo S, Fairbrother GL, et al. (2015) Evaluating intra- and inter-individual variation in the human placental transcriptome. Genome Biol 16: 54. doi: 10.1186/s13059-015-0627-z PMID: 25887593