JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 28, Number 6, 2021
Mary Ann Liebert, Inc.
Pp. 527–559
DOI: 10.1089/cmb.2020.0032

Toward an Information Theory of Quantitative Genetics

DAVID J. GALAS, JAMES KUNERT-GRAF, LISA UECHI,* and NIKITA A. SAKHANENKO

ABSTRACT

Quantitative genetics has evolved dramatically in the past century, and the proliferation of genetic data, in quantity as well as type, enables the characterization of complex interactions and mechanisms beyond the scope of its theoretical foundations. In this article, we argue that revisiting the framework for analysis is important and we begin to lay the foundations of an alternative formulation of quantitative genetics based on information theory. Information theory can provide sensitive and unbiased measures of statistical dependencies among variables, and it provides a natural mathematical language for an alternative view of quantitative genetics. In the previous work, we examined the information content of discrete functions and applied this approach and methods to the analysis of genetic data. In this article, we present a framework built around a set of relationships that both unifies the information measures for the discrete functions and uses them to express key quantitative genetic relationships. Information theory measures of variable interdependency are used to identify significant interactions, and a general approach is described for inferring functional relationships in genotype and phenotype data. We present information-based measures of the genetic quantities: penetrance, heritability, and degrees of statistical epistasis. Our scope here includes the consideration of both two- and three-variable dependencies and independently segregating variants, which captures additive effects, genetic interactions, and two-phenotype pleiotropy. This formalism and the theoretical approach naturally apply to higher multivariable interactions and complex dependencies, and can be adapted to account for population structure, linkage, and nonrandomly segregating markers. This article thus focuses on presenting the initial groundwork for a full formulation of quantitative genetics based on information theory.

Keywords: entropy, epistasis, genetics, information theory.

1. INTRODUCTION

THE CRITICAL QUESTIONS for understanding a genetic system, its functions, structure, and complexity, lie in the actual dependencies among the system's variables, both the phenotypes and genotypes, as well as external factors. The phenotypes, of course, can range from highly specific cellular or molecular measures to broader-, functional-, and organismal-level phenotypes. The information architecture of the genetic system's

Pacific Northwest Research Institute, Seattle, Washington, USA.
*Current address: Beckman Research Institute of City of Hope, Los Angeles, California, USA.

variables is at the heart of the dependency problem, and the difficulty of determining this architecture from data is significant for truly complex systems, which well describes many important genetic problems. These problems are inherent in the challenges of the past concerning the genetic explanation of complex traits, the notion of missing heritability, and the complex effects of gene interaction.

Quantitative genetics has evolved substantially over the 100 years since Fisher and Wright laid its foundations in these articles (Fisher, 1918; Wright, 1926), for example. It has been pointed out repeatedly, however, that while their methods were powerful and innovative, there are some problems with the general approach and the tacit assumptions inherent in them (Nelson et al., 2013). It is not that the classical methods are not correct and powerful, but rather that there are unanticipated subtleties and tacit assumptions that are not recognized. The proliferation of new data types calls for additional approaches and different mathematical descriptions, and since the logic of using the classical variance methods to infer genetic architecture is flawed (Huang and Mackay, 2016), new approaches are needed for this reason as well.

Nelson et al. (2013) have argued effectively that the Fisherian paradigm has reached its limits in the ability to deal with complex traits and modern genetic data. Their summary, ''… many of the current tools are adaptations of methods designed during the early days of quantitative genetics. The present analysis paradigm in quantitative genetics is at its limits in regard to unraveling complex traits and it is necessary to re-evaluate the direction that genetic research is taking for the field to realize its full potential,'' is a clear call for new quantitative approaches. It is also true that despite the innovative statistical approach in 1918, the Fisherian methods have often been misunderstood and/or misused in present quantitative genetics.

Huang and Mackay (2016) have pointed out and clearly made the case that the genetic architecture of quantitative traits simply cannot be inferred from variance component analysis, which has been applied for that purpose in many studies over the years. The logic of this use of variance analysis is simply wrong because the underlying assumptions that would allow such inference do not generally hold. It is clear that genetic interactions, called epistasis, in a common use of this term, have been implicated as essential for understanding complex traits (Phillips, 2008; Eichler et al., 2010; Gilbert-Diamond and Moore, 2011; Hill et al., 2014; Mackay, 2014). Although it has also been challenged as being unimportant in evolution (Crow, 2010), this seems unlikely to us. Recent results have, on the contrary, strongly supported the importance of interactions in understanding complex traits (Phillips and Johnson, 1998; Gregersen et al., 2006; Wiltshire et al., 2006; Coutinho et al., 2007; Phillips, 2008), including those in humans, and it would seem that evolution cannot escape such an influence. In addition, quantitative inference of interacting loci will likely be important for understanding polygenic risk scores, which are currently being generated using noninteracting largely additive models.

Here we propose that information theory can provide the foundations of a new approach to quantitative genetics, which focuses on the information content of the genome and the advantages of information theory, and we begin the process of building that foundation with this article. It is not our position that present methods are faulty, but rather that it is likely that establishing a new approach and formulation will reveal new insights and provide new methodologies because of the fundamentally different viewpoint. For example, the ability to detect two-locus dependencies without significant single-locus dependence extends the analysis power beyond the Genome-Wide Association Studies (GWAS) method. This extension is a natural feature of the information theory formulation.

The application of information theory to genetic problems actually has a long history. It begins with the surprising fact that Shannon (1948), the architect of information theory, actually wrote his PhD thesis in 1940 on ''a new algebra of genetics'' (Shannon, 1940; Crow, 2001), which addressed some key issues in population genetics at the time. In later work, issues relating to the relationship between evolution and the statistics of population genetics were tackled using concepts from Shannon's information theory (Moran, 1961; Watterson, 1962; Frieden et al., 2001).

Information theory, while originally directed at understanding communications quantitatively, has been very effective well outside of this original domain and has been applied widely to physical, biological, and chemical problems, and to other fields (Jaynes, 1957, 2005; Galas et al., 2014). In almost all scientific domains, the problem of inferring the quantitative dependencies among measurable variables, and even causal relationships, is the central problem, and the information measures, functionals of probability distributions, have been shown to be powerful tools in these problems of inference. We have previously shown how information theory methods can be used to analyze complex data, and have also shown how genetic data are amenable to some such applications (Galas et al., 2014; Sakhanenko and Galas, 2015; Sakhanenko et al., 2017). Here we extend both the formulation of the relationships and methods and their interpretation and recast the theory into a more comprehensive description of quantitative genetics. While we take only the first few steps here toward a full information theory of complex genetics, we show how this approach forms a fruitful way to describe the complex genetic architecture of a system. Specifically, we describe familiar concepts such as gene interaction, pleiotropy, penetrance, degree of epistasis, and heritability in terms of information theory.

Our concept of the information architecture of a system derives primarily from the idea of using information measures to define the levels of dependencies among variables. Information theory, being model-free, is broadly applied to extracting statistical properties from the data, which are in turn determined by the joint probability distributions of the variables.

Information measures have the advantage of being completely agnostic of any models or prior assumptions affecting dependencies, unlike many commonly used methods in genetics, particularly including correlation methods. This model-free character allows the data to fully drive the conclusions. These methods also reduce the sensitivity of the measures to small variations in the data, and to the limitation of small sample sizes. Thus, we argue here that the application of information theory to genetics can provide a powerful approach to deciphering the structure of complex genetic systems and to extracting their information architecture, which is distinct from the genetic or model architecture. This article advances our previous work in which we defined an information landscape (Sakhanenko et al., 2017) and illustrated the use of discrete functions and noise on this landscape to analyze genetic data. Here we focus on specifically elucidating the relationships of three-variable dependencies and complete this picture by providing a way to extract the specific functional nature of dependencies for variables whose dependency has been detected and measured.

General as it is, the application of information theory to any specific area carries with it certain assumptions and premises, which need be made explicit. The principal caveats that must be addressed are these. The idea that the statistical inferences from the data reflect the subtle features of variable dependency assumes that the sampling issues and density of data represent these features in sufficient detail for information methods to make reliable estimates of the fundamental quantities, the entropies. In actual use this is often a rough approximation only and the approximation must be explicitly quantitated and its meaning acknowledged. We discuss this question later in the article and for the purpose of explication initially simply assume for the moment that the data set is large enough to be fully reflective of the underlying relationships.

It is also clear that by its nature, information theory is inadequate to fully represent some distinctions among certain distributions. There are indeed distinctly different distributions with identical information measures. The mapping of probability distributions of variables into information measures is decidedly many-to-one. There are therefore several models and architectures that may have the same sets of measures. Another caveat depends on the question of how many variables participate in synergistic dependencies in a complex system since the number must be carefully controlled in any practical application because of statistical and computational limits (Galas et al., 2017). While the method is entirely general, we limit ourselves in this article to considering two- and three-variable dependencies only. This is sufficient to demonstrate the formulation and to illustrate its usefulness, and the power of the three-variable method is amply demonstrated.

To make the formulation more self-contained, we add a short primer on the key information theory quantities. First and foremost is the definition of Shannon's entropy. For m possible states of a variable, X, $\{x_i\}$, where the probability of a sample or subject i having a value of x_i is p_i, the entropy of the variable X in this data set is $H(X) = -\sum_{i=1}^{m} p_i log p_i$. The joint entropy of two variables X and Y is defined in the same way where the possible states are those of the pairs $\{(x_i, y_j)\}$. The conditional entropies are obtained from the Shannon formula by simply using the appropriate conditional probabilities.

An important measure that assesses the information in one variable about another is the mutual information. For two variables, X and Y, this is denoted $I(X,Y)$, and is defined as $I(X, Y) = H(X) + H(Y) - H(X, Y)$. If the two variables are independent of one another this is zero, as expected, since the joint entropy in this case is simply the sum of the two single entropies. The joint entropy can be extended to three variables by simply using the distribution of values of triplets of the variables, which are also obtained from the data in practice. All of the information measures used here can be expressed as sums and differences of entropies. In Appendix E, we briefly address the important issue of estimation of entropies from data, the accuracy of which depends most sensitively on the amount of data available and the range of variable values considered. The errors in our calculation of entropies can be estimated, and must be kept in mind, but we rely here on the small variable alphabet sizes, and the large number of data samples to keep these small.

The symmetries of the relationships among the information functionals are surprisingly simple, but also subtle. The multiple measures of information theory have strikingly symmetric relations and number of symmetries that we have previously reported (McGill, 1954; Bell, 2003; Galas et al., 2010; Sakhanenko and Galas, 2019). The symmetries all derive from the fact that all information measures are specific linear combinations of joint entropies, such as the mutual information, organized by lattices whose partial order is determined by inclusion of variable subsets.

In addition, there are a number of problems that can be fully analyzed for discrete functions, which are the most common manifestations of the variables we deal with in data analysis. By this we mean that the dependent variables in a complex system can be viewed as functions of one another, and the discrete values of the data can therefore be viewed as reflecting these discrete functions. While real genetic data have various levels of probabilistic determinants and ''noise,'' much of the character of the dependency can be represented by multivalued discrete functions, which are mixed with various levels of ''noise'' to describe the realistic intervariable dependencies. This gives us a distinct mathematical advantage since, in principle, we can characterize the properties of all possible discrete functions with finite alphabets. We examine here the properties of discrete functions, and their information architecture and relationships show in detail how functions can be classified, and examine the extension of this analysis to include probability density functions that result from adding ''noise'' or subtracting determinism from the discrete functions.

2. OVERVIEW OF FORMALISM

The complexity of genetics arises not only from the interactive functions encoded in the genome, and the range and complexity of phenotypes, but also from the structures of study populations and inheritance patterns in complex pedigrees. In this article, while recognizing the important effects of population structures on quantitative genetic measures, we defer addressing these important issues so that we can restrict our considerations here to large, randomly mating populations, described as *panmictic,* recognizing that no natural population is fully panmictic, and few artificial, experimental populations are panmictic in practice. We will consider population structure issues in a later article.

The basic components of the formalism presented here are summarized in these five points:

1. The information measure we call the *symmetric delta* (Galas et al., 2014), as shown in Section 4, is used to detect the dependence of subsets of loci with phenotypes in the data. In this article, we consider pairwise and three-way dependencies.
2. The general relation between genetic loci and phenotypes is embodied in a discrete valued *loci/phenotype array*: $f(X_1,\ X_2,\ \dots\ X_n)$, where $\{X_i\}$ is the set of n genetic loci and the function determines the phenotype. This is identical in two dimensions to what geneticists often embody in a matrix connecting three variables, called a ''gene/phenotype table.'' We limit ourselves to one or two genetic variables (loci) here. Without loss of generality we could include multiple phenotype variables as well.
3. The essential ''noise'' distributions, when added to these arrays, form the genotype/phenotype arrays, which describe the phenotype in terms of loci, noise, and penetrance

$$G(X_1,\ X_2,\ \dots X_n;\ p) = pf(X_1,\ X_2, X_3\ \dots\ X_n) + (1-p)\varepsilon$$

   where ε is a ''noise'' function, and (1-p) is the noise level (p is the penetrance). The noise can be assumed to follow a particular structure (e.g., uniform random noise).

4. The arrays *f* and *G* for tuples of variables with significant dependence are inferred from the data using relatively simple algorithms.
5. These arrays are then used to calculate penetrance, heritability, gene interactions, and pleiotropy.

This article is organized as follows. We first present the basic discrete function expression of genetics, gene/phenotype tables, assuming full genetic dependence (with no ''noise''), and then review the basics of the information measures previously introduced (Galas et al., 2014, 2017; Sakhanenko and Galas, 2015; Sakhanenko et al., 2017). We then describe some specifics of three-variable dependencies and the symmetries that their information measures exhibit (Sakhanenko and Galas, 2019). We review the information landscape notion we previously proposed (Sakhanenko et al., 2017) and extend it to a more general form.

Introducing a formal array structure for extending the information landscape allows us to systematically handle all probability distributions, which are essential for the introduction of ''noise,'' for arbitrary size alphabets (possible discrete values of variables). This formulation shows that the information content of the discrete functions is strongly dependent on both the alphabet size and the symmetries of the functions. This rich area is partially explored here but provides us some initial insights and a flexible set of theoretical tools with which to characterize complex genetic systems. We then define a set of transformations that map the three-variable functions into a two-variable function space and allow us to greatly simplify the identification of the functional structure of the inferred dependencies.

We discuss the implications of these results and tools for the analysis of genetic data using information-based methods, and describe, in addition to penetrance, the genetic notions of gene interaction, pleiotropy, and heritability in terms of information theory measures. Finally, we apply our methods to some real yeast data and discuss the analysis of complex genetic data (Bloom et al., 2015).

3. DISCRETE FUNCTIONS AND GENETICS

The classic genotype/phenotype table for two loci can be usefully considered a discrete function where the phenotype variable, *Z*, is expressed as a function of the two genotype variables, *X* and *Y*. Diploid binary alleles, variants for *X* and *Y* are, of course, three-valued, and haploid binary variants are two-valued. While the phenotype alphabet can be any size, in principle, we also use three values for the phenotype alphabet (0,1,2). The alphabet can certainly be expanded to include more than binary allele variants, but for simplicity we do not consider these in this article. Often a two-valued variable is sufficient to effectively describe a phenotype, but quantitative phenotypes require larger alphabets. These tables are similar to the Punnett square in classical genetics. Consider the discrete functions where all three variables, *X*, *Y*, and *Z* are three-valued, and $Z=f(X,Y)$, with *X* and *Y* independent. Each of the functional relationships can be represented by a 3-by-3 table. Table 1, for example, shows three functions that can be seen as tables for logical AND, logical exclusive OR (XOR), and equality (EQ) functions, extended to three variables. In these tables the genotypes are encoded as follows: 0 = homozygous major alleles, 1 = heterozygote, and 2 = homozygous minor alleles.

These discrete functions describe the phenotype as a function of the two genetic loci. The EQ function is a degenerate case for which *Z* is only a function of *X*. The general scheme can also be thought of as implementing a three-valued logic. We can call the function defined by the left-most table (Table 1) an extended AND, because the lower value of the two arguments, X and Y, dominates, as in binary logic. Since there may be phenotype determinants other than these two loci, as described above, we will generally need a more complex function to describe the phenotype in a population. A three-variable array, which we consider in detail in a later section, can embody this complexity. The discrete function is generally modified by the random effects of both unknown and stochastic effects, the ''noise'' represented by a random function, and the penetrance, the degree of determination of the phenotype by the genetic variables.

The diploid case considered here is, of course, most commonly encountered in genetics of mammals, but the haploid case is not unknown in genetic data, for example, in the case of recombinant inbred populations and organisms that can grow and divide as haploids. We apply our methods to an example of a haploid case in data from *Saccharomyces cerevisiae*. In the case of haploid genetics, the alphabet of values for the genetic variables is binary so that the genotype/phenotype table is 2 × 2 and the logic is essentially Boolean. This simplification can be very useful for practical calculations, as we discuss later.

Table 1. Examples of Extended (3-by-3) Tables Defining Three-Valued Versions of Genetic Functions

			X					X					X	
	AND	0	1	2		XOR	0	1	2		EQ	0	1	2
Y	0	0	0	0	Y	0	0	1	2	Y	0	0	1	2
	1	0	1	1		1	1	2	0		1	0	1	2
	2	0	1	2		2	2	0	1		2	0	1	2

From left to right these correspond to logical AND, XOR, and EQ functions. These functions can be represented in a linear notation (by reading the tables left-to-right and top-to-bottom) as 000011012, 012120201, and 000111222, respectively.

EQ, equality; XOR, exclusive OR.

4. ELEMENTS OF THE THEORY

4.1. Genetic dependence relations

We begin by reviewing some definitions and previous results, and then introduce extensions of these relations. The first important point is that mutual information, an inherently pairwise measure, is unable by itself to capture the full information in dependencies. Full representation requires many variable subsets, but even for three-variable tuples considered here, mutual information is insufficient and requires additional measures to fully characterize the dependence among three variables. As has been pointed out before, a clear example is the XOR relationship for any size of alphabet (Sakhanenko et al., 2017). For the binary alphabet, three-variable case, it is evident that the mutual informations between all pairs of variables for this function vanish. We have demonstrated that the ternary XOR-like functions (Table 1) also exhibit this property (Sakhanenko et al., 2017). It is also true for any size alphabet and is reflective of the symmetry of the dependencies.

Even the interdependency of two variables has a surprising level of complexity in the ways it can be expressed. Mutual information has several equivalent mathematical expressions. The most common form is as a difference of entropies, as described in the introduction. In terms of the conditional entropies we also have these symmetric expressions for mutual information.

$$I(X, Y) = H(X) - H(X|Y) \text{ or } I(X, Y) = H(Y) - H(Y|X) \tag{1}$$

An important information measure is a generalization of mutual information for multiple variables, called the interaction information, or coinformation (McGill, 1954; Bell, 2003). For n variables this is defined by the recursion relation

$$I(X_1, X_2, X_3 \ldots X_n) = I(X_1, X_2, X_3 \ldots X_{n-1}) - I(X_1, X_2, X_3 \ldots X_{n-1}|X_n) \tag{2}$$

This measure can also be expressed by the sums and differences of joint entropies of the full set of variables $\nu = \{X_1, \ldots X_n\}$, (represented by Möbius function for the lattice of subsets, τ, in this formula)

$$I(\nu) = \sum\nolimits_{\tau \subset \nu} (-1)^{|\tau|+1} H(\tau) \tag{3}$$

For three variables, the interaction information is simply expressed in terms of the entropies.

$$I(X_1, X_2, X_3) = H(X_1) + H(X_2) + H(X_3) - H(X_1, X_2) - H(X_1, X_3) - H(X_2, X_3) + H(X_1, X_2, X_3) \tag{4}$$

We define the *differential* interaction information (delta) as the change in the interaction information that occurs when we add another variable to the set. In general, where $X_m \notin \nu_m$ and $\nu_m \cup X_m = \nu$, the differential interaction information is defined as

$$\Delta(\nu_m; X_m) = I(\nu) - I(\nu_m) = \sum\nolimits_{\tau_m \subseteq V | X_m \in \tau_m} (-1)^{|\tau_m|+1} H(\tau_m) \tag{5}$$

The general measure of the fully collective dependence among all variables, the symmetric Delta, $\bar{\Delta}$, is defined as the product of deltas:

$$\bar{\Delta}_m = \bar{\Delta}(\nu_m) \equiv \prod\nolimits_{i=1}^{m} \Delta(\nu_i; X_i) \tag{6}$$

This is the measure we have previously proposed and used for measuring collective dependence of a set of variables. For three variables, the differential interaction informations (the ''deltas'') can be obtained by permutation of the variables.

$$\Delta(X, Y; Z) = I(X, Y, Z) - I(X, Y) = -I(X, Y|Z) = H(Z) - H(X, Z) - H(Y, Z) + H(X, Y, Z)$$

and the symmetric delta for three variables is

$$\bar{\Delta}(X, Y, Z) = \Delta(X, Y; Z) \cdot \Delta(Z, Y; X) \cdot \Delta(X, Z; Y) \tag{7}$$

We can also usefully expand the n variable joint entropy $H(X_1, \ldots X_n)$ into a sum of terms, each of which depends on the number of variables, using the Möbius inversion (Sakhanenko and Galas, 2019). This gives us an expression for the entropy as a sum over interaction informations over all possible subsets of variables. This approach generates a series of approximations in the number of variables considered, and represents a practical, general, and systematic way forward in the genetic formalism for more than three total variables, in that it provides the appropriate approximation for each limiting assumption. We will illustrate and use this approach in future work.

4.1.1. Multi-information as total dependence. Another important information measure that we will use in several ways is the multi-information for n variables (originally defined and called "total correlation," by Watanabe (1960) and discussed and used by many others (Ting, 1962; Han, 1980)). It is defined as the difference between the sum of entropies of each variable separately and the joint entropy of all the variables together:

$$\Omega(X_1, \ldots X_n) \equiv \sum_{i=1}^{n} H(X_i) - H(X_1, \ldots X_n) \tag{8}$$

The multi-information is essentially the collective measure of all dependencies among the n variables; that is, the sum of dependencies for all possible subsets of variables. It is zero only when all the variables are independent, so it does not distinguish among the orders of dependency. This stands in contrast to the symmetric delta, which is the measure of the full synergistic dependency of all the n variables together. It is zero when any one of the variables is independent of the others. Since the multi-information deals with dependence of all possible subsets, and the symmetric delta deals with dependence of the entire set, they are like bookends of the dependency measures. As shown in the next section, the multi-information is a key element in the quantitative relationships we use in this formalism.

4.1.2. Three-variable dependencies. While the restriction to pairwise dependency analysis is equivalent in concept to classical association studies in genetics, sufficient for some problems, the detection of even three-variable dependencies can add much to the power of the analysis and is essential for any genetic system that involves pleiotropy or gene interaction. Note that pleiotropy is defined as the dependence of one genetic variant variable and two phenotypic variables. We focus in this section on understanding the key relations for systems at the three-variable level. The relations among the three-variable information measures are simple, but subtle, and illustrate the strong symmetries inherent to the information measures. Furthermore, it is useful to examine carefully the bounds on their values. However, first, a few more preliminaries.

From here on we use a simplified notation, where the three variables are labeled by integers: $X \to 1$, $Y \to 2, Z \to 3$. Wherever the meaning is clear, we abbreviate using these labels within a subscript; for example, $H(X, Y, Z) \to H_{123}$ and $I(X, Y) \to I_{12}$. The relations between the mutual informations and the multi-information, and the deltas (where we define the notation $\Delta_1 \equiv I(2, 3|1)$, $\Delta_2 \equiv I(1, 3|2)$, $\Delta_3 \equiv I(1, 2|3)$) are provided by these equations:

$$\begin{aligned} \Delta_1 &= \Omega_{123} - I_{12} - I_{13} \\ \Delta_2 &= \Omega_{123} - I_{12} - I_{23} \\ \Delta_3 &= \Omega_{123} - I_{13} - I_{23} \end{aligned} \tag{9a}$$

We derived these previously (Sakhanenko et al., 2017), but they are easily shown to be true by simply expressing the measures Δ, Ω, and I in terms of sums and differences of entropies. Since Ω always refers to all three variables, we can drop the subscripts for this quantity without ambiguity in most cases. For two variables only, of course, $\Omega = I_{ij}$, the mutual information. It can easily be seen that these equations are symmetric in the variables, and the only asymmetry arises from the differences among these terms. The above relations for three-variable dependencies can, of course, also be formulated conveniently in matrix form, which is shown in Appendix C. This matrix equation may be a useful tool for further exploration of three-way dependence symmetries.

For genetic data, where X and Y are independently segregating genetic loci, valid for panmictic populations, and Z is the phenotype variable, the three mutual informations in Equation 9a become two since $I(X, Y) = I_{12} = 0$. The assumption of independently segregating variants is essentially equivalent to assuming linkage equilibrium. In this case there are only three relevant measures in the set of relations (Equation 9a), Ω, I_{13}, and I_{23}, and the relationship is significantly simplified.

$$\begin{aligned}\Delta_1 &= \Omega - I_{13}\\ \Delta_2 &= \Omega - I_{23}\\ \Delta_3 &= \Omega - I_{13} - I_{23}\end{aligned} \tag{9b}$$

We can normalize Equation 9b by dividing through by Ω, the total of all dependencies, as long as there is some dependency so that $\Omega > 0$. We get the normalized delta coordinates (only for the case of $I_{12} = 0$), which were the coordinates used in Sakhanenko et al. (2017) to define the geometry of the information landscape. The coordinates of the information landscape are these:

$$\begin{aligned}\delta_1 &= 1 - \mu_{13}\\ \delta_2 &= 1 - \mu_{23}\\ \delta_3 &= 1 - \mu_{13} - \mu_{23}\end{aligned} \tag{10}$$

where the δ's are the normalized Δ's, and the are the μ's normalized mutual informations.

We can rearrange the above equations into a simple relation for δ_3 as a function of δ_1 and δ_2 :

$$\delta_3 = \delta_1 + \delta_2 - 1 \tag{11}$$

The condition for δ_3 to be nonzero then is $\delta_1 + \delta_2 > 1$. This is one side of the line defined by $\delta_1 + \delta_2 = 1$. Let us look more closely at the constraints on δ_1 and δ_2 imposed by $I_{12} = 0$. If we look at the three-dimensional (3D) space defined by the three δ's, which is what we call the information landscape, we can see that we have three coordinates and one linear constraint that thereby defines a two-dimensional plane. One natural question regarding bounds of the landscape is whether negative coordinates are possible. The answer is that they are not. The key inequalities that bound these quantities are intuitive and elementary, but still not entirely obvious and we state them explicitly and present the proofs in Appendix A.

The interaction information, I_{123} , is defined in terms of the entropies (Bell, 2003; MeGill, 1954) as

$$I_{123} = H_1 + H_2 + H_3 - H_{12} - H_{13} - H_{23} + H_{123}$$

and by the definitions of mutual information we have

$$\Omega = I_{12} + I_{13} + I_{23} - I_{123} \tag{12}$$

Notice that if $I_{12} = 0$, by proposition 2 this expression implies that $I_{123} \leq 0$.

A few more points about dependencies among genetic variables are in order here. Equation 9b applies when I_{12} is strictly zero, however, I_{12} may as well be nonzero in real data, because of disequilibrium or because of noise in the data, including sampling-induced fluctuations. We will deal with the linkage disequilibrium (LD) issue in a future article, but it is important to note that even in the presence of LD, the symmetric delta represents the full interaction score for any triplet, including the contribution due to LD. A significant problem to be discussed in a future publication is that it is more difficult in this case to extract the quantitative score for the strictly three-way component (we call this the epistatic component). The potential entanglement of epistasis and LD, which is often overlooked in genetic analyses, is at the heart of this issue.

There are many ways of expressing the set of relationships described above for three variables. For example, Equation 9 leads directly to the expression for the multi-information as in Equation 14. Since it is clear that if the dependencies are pairwise, and $I_{12} = 0$, then the mutual informations contain all the dependence, in which case $\Omega = I_{13} + I_{23}$. Thus, in this case for three-way dependence, we can ascribe the epistatic component (three-way) to the value of $-I_{123}$ (the minus sign comes from our sign convention above). In other words, in the case of linkage equilibrium, the interaction information is the epistatic dependence measure. This is a useful way to decompose the multi-information. This relation for the triplet dependencies is illustrated in Figure 1. We emphasize again that the epistatic component is $-I_{123}$ only when $I_{12} = 0$. In the general case, $-I_{123}$ is equal to the epistatic component minus the information shared by 1 and 2 affecting 3. The above equations allow us to define several important limiting conditions. This is further illustrated in Figure 2. We can summarize these constraints on the basic measures and their implications or interpretations simply, and this is presented in Table 2.

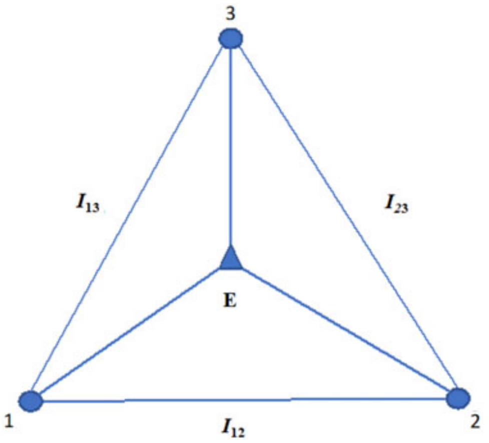


FIG. 1. Three-variable dependencies that make up the multi-information or total correlation (we adopt the convention here that X is 1, Y is 2, and Z is 3). The lines represent the components of dependence among the variables (small circles) as in the above equation, where the epistatic component is represented by the lines emanating from the triangle. The epistatic component is $E=-I_{123}+S$.

4.1.3. The components of genetic dependency and their measures. The genetic architecture of a phenotype is determined by the dependencies among the genetic variables and the phenotype variable. The application of the information formalism can, however, be rather subtle and care must be taken in its interpretation. In this section, we define the problem in a bit more detail and make the specific connections between information theory quantities and genetic quantities.

The dependencies of phenotypes on more than one genetic locus define what we mean by genetic interactions and consist of a wide, but finite, range of possible forms of interactions. These effects have been recognized for 110 years when William Bateson proposed this as an explanation for deviations from simple Mendelian ratios. If a phenotype is dependent on two loci, each exclusively in a pairwise manner, we call this effect additive, and distinguish it from what in the usual terminology is called an epistatic effect or interaction. Fisher called this statistical epistasis *epistacy* and attributed the deviations from additivity to his linear statistical model. Many modern authors have argued and provided evidence that gene/gene interactions are rather common (Gilbert-Diamond and Moore, 2011). The most common way to deal with these interactions quantitatively, however, has been to use regression methods (Lstiburek et al., 2018), and more recently, other machine learning tools. In all these cases, however, the starting loci are most often those identified by GWAS or some pairwise method, which will then miss those loci that are invisible to pairwise methods.

Quantitating ‘‘gene interaction,’’ that is, measuring the amount of the phenotype that depends on the combined markers, can be done naturally with the measures defined here. We need to be precise, however, in defining what we mean by gene interaction, and we need to distinguish additive effects from epistatic interactions, the former being strictly pairwise, the latter not including any pairwise effects. Again, we are here assuming independently segregating variants and $I_{12}=0$.

If the genetic variant variables are X and Y, and the phenotype variable is Z, we consider all possible three-variable dependencies, as in Figure 1. In this general three-variable case, we can quantitate the

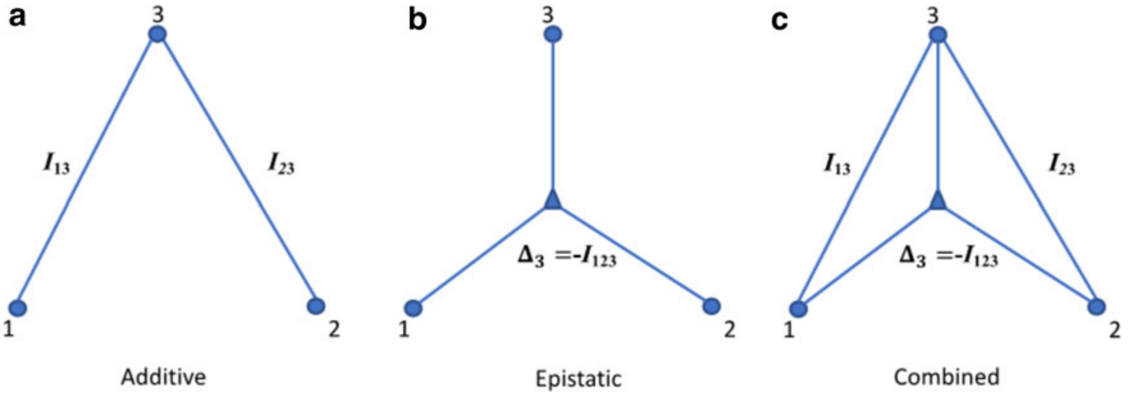


FIG. 2. Independent segregation interaction relationships. The genetic contributions of 1 and 2 to the phenotype, 3, illustrating the distinction between the additive **(a)** and epistatic **(b)** effects within a relationship with a combined effect **(c)**.

Table 2. Several Limiting Constraints on the Information Relations, with Their Interpretations or Consequences

Constraint	*Consequence*
$I_{12}=0$	Independent segregation of loci implies linkage equilibrium
$\Delta_3=0$	Pairwise dependencies only, $\Omega=I_{13}+I_{23}$; $\Delta_1=I_{23}-I_{12}$; $\Delta_2=I_{13}-I_{12}$
$I_{12}=I_{13}=I_{23}=0$	Three-way dependence only, $\Delta_3=\Omega$; $\Delta_1=\Delta_2=\Omega$

Keep in mind that these rules apply strictly only to the discrete functions without noise.

information contribution of X and Y to the determination of Z by the mutual information between Z and the joint X,Y variables, $I(Z,(X,Y))$. Using the mutual information chain rule

$$I(Z,(X,Y))=I(Z,Y)+I(Z,X|Y) \tag{13}$$

and identifying $I(Z,X|Y)=\Delta_Y$ and using Equation 9a, we have simply

$$I(Z,(X,Y))=\Omega-I(X,Y) \tag{14}$$

In the case of independent segregation of markers, where $I(X,Y)=0$, this becomes $I(Z,(X,Y))=\Omega$, as expected since in the absence of shared information between X and Y, the mutual information $I(Z,(X,Y))$ describes the full extant dependence. As shown in the previous section, the decomposition of the information contributions becomes simple in this case. In Figure 2, we illustrate the nature of the dependencies.

We wish to emphasize that the relationships present here permit the decomposition of the information structure, the dependencies, of the variables. It is important that the dependency can be decomposed and we can determine what fraction of the dependence is pairwise and what fraction is three-way dependent (synergistic or epistatic). These fractions can be derived simply from the equation $\Delta_3=\Omega_{123}-I_{13}-I_{23}$ provided that $I_{12}=0$. Since the pairwise dependence of the phenotype on the two loci is the sum of the mutual informations, I_{13} and I_{23}, and the total dependence is Ω_{123}, the three-way dependence is given by their difference, Δ_3. The fractional dependencies are then simply the ratios

$$\text{Pairwise fraction}=F_p=\frac{I_{13}+I_{23}}{\Omega_{123}},\ \text{Epistatic fraction}=F_e=\frac{\Delta_3}{\Omega_{123}} \tag{15}$$

If the genetic variables are 1 and 2, and the phenotype is 3, these fractions represent the pairwise additive contributions of 1 and 2 to the phenotype, F_p, and the nonadditive, or epistatic, contribution, F_e.

The Equations 14 and 15 apply in this case and the separation of the additive and nonadditive, or epistatic, effects is clear.[†] We will address the more complex case of nonzero LD and related effects in a future article.

The epistatic interaction in the case of no disequilibrium is measured entirely by Δ_3. This is also rather intuitive since the multi-information, Ω_{123}, quantitates the total dependence and the mutual information quantitates the pairwise dependencies between each variant and the phenotype. Thus, their difference measures epistatic gene interaction.

$$\textit{Equilibrium epistasis measure}=\Delta_3=\Omega_{123}-I_{13}-I_{23} \tag{16}$$

There is another kind of three-variable dependence that is important in genetics. A single genetic locus affecting two distinct phenotypes, which is called *pleiotropy*, can be described by the general equations, but the limiting constraint of independent segregation, which makes the mutual information between variants vanish, does not apply in this case. The analogous constraint, however, is that the mutual information between phenotypes vanishes. We consider pleiotropy briefly in our discussion of the yeast data. Full pleiotropic analysis can be rather complex, however. For example, unlike for two genetic variables, we cannot easily understand what it means to decompose the information contributions of one genetic variant

[†] If $I_{12}\neq 0$ the situation is more complex, and distinctions need to be made between the redundant information provided by X and Y, the unique information provided, the synergistic information, and the quantities indicated in Figure 2 do not apply. This is essentially the information decomposition problem, which has no universally accepted method of computation (Bertschinger et al., 2012). We deal with this in a future publication.

and phenotype on a second phenotypic variable. The potential complexities are both interesting and significant and will be considered in future work.

4.2. Information theoretic relations and symmetries

When two loci (X,Y) are involved in determining a phenotype, Z, we can represent the relation as a genotype/phenotype matrix. These three-variable matrices have discrete values and thus are discrete functions of two variables, $Z(X,Y)$. We have shown that the 3D information landscape, defined by the three normalized deltas from Equation 10, is a plane when $I_{12}=0$, and under this condition, all discrete functions lie on this plane.

4.2.1. Discrete functions. There are several possible ways of defining the information content of discrete functions, and discrete functions are a useful way to characterize quantitative genetic relations. The usual genotype/phenotype tables for two genetic loci used in classical genetics are just this kind of discrete function, and therefore, the information in these functions is the key to quantitative analysis. Here we define the inherent information as the measures calculated from the probabilities inferred directly from the function. Note that all discrete functions map into distributions, but not all distributions are discrete functions. Information measures (all are linear combinations of joint entropies) are functionals of distributions. However, the finite size of the set of discrete functions (for a given set of alphabet sizes) and the (infinite) size of the set of all possible distributions are incommensurate. There are a finite number of discrete functions for any finite number of variables and alphabets, but there is an infinity of distributions for any finite number of variables and alphabets. The addition of ''noise'' to the discrete functions generates an infinite range of distributions. As seen, this is a key consideration in quantitative genetics. The ''noise'' determines the penetrance of the genetic dependence on the discrete function.

As shown in Sakhanenko et al. (2017), we can map all the discrete functions onto the information plane (e.g., there are 19,683 functions on this plane for the 3×3 case). When the information measures are calculated for the 3×3 functions and plotted in the plane, they form simple rectilinear patterns for each value of Ω. The positions of all function families (those functions with identical normalized delta coordinates) are shown in Figure 3a.

In Figure 3b, the families are shown in different panels for different values of Ω, total dependence. Even though all functions in a family have the same delta coordinates, not all of the functions in a family need have the same value of Ω. Notice the symmetry in the triangular plane that results from the exchange of X and Y.

In the case of haploid genetics, the information plane for three variables shows a similar geometric symmetry, but with many fewer functions. Many published yeast genetic data sets are haploid, including the data we have analyzed here to demonstrate our methods (Bloom et al., 2015). Haploid genetic state variants are binary and since there are N^4 discrete functions, where N is the alphabet size for the phenotype, Z, this can lead to a significant simplification of the information landscape for binary phenotypes. For $N=2$, there are only 16 functions in all, but as N increases from 3 to 5, the number of functions grows rapidly, and there are 8 families of functions, each family having identical information coordinates. As the phenotype alphabet size, N, increases past 5, the number of families stays the same even as the number of functions grows rapidly. In Figure 4, the information plane and the families are shown.

Finally, since we propose to use the symmetric delta of Equation 7a to find three-way dependencies, it is natural to ask if the total dependence were equal for multiple triplets, which discrete function would maximize the symmetric delta? As we show in Appendix B, the answer is that it is the triplet with no pairwise dependence and only three-way dependence, the XOR-like functions. This is interesting in several respects, first because having the product of the conditional mutual informations maximal for discrete functions where pairwise dependence vanishes seems unexpected, but more importantly, because a distribution of triplet symmetric delta scores, maximizing at the XOR-like functions, is a very useful indicator of specific functional dependence.

5. INFERRING THE FUNCTIONAL DEPENDENCE OF PHENOTYPE ON GENOTYPE

5.1. The meaning of ''noise'' in genetic data

There will always be ''noise'' in the data, which arises from two classes of sources, unknown variables and stochastic processes inherent to the biology and data acquisition processes. The word is in quotes here

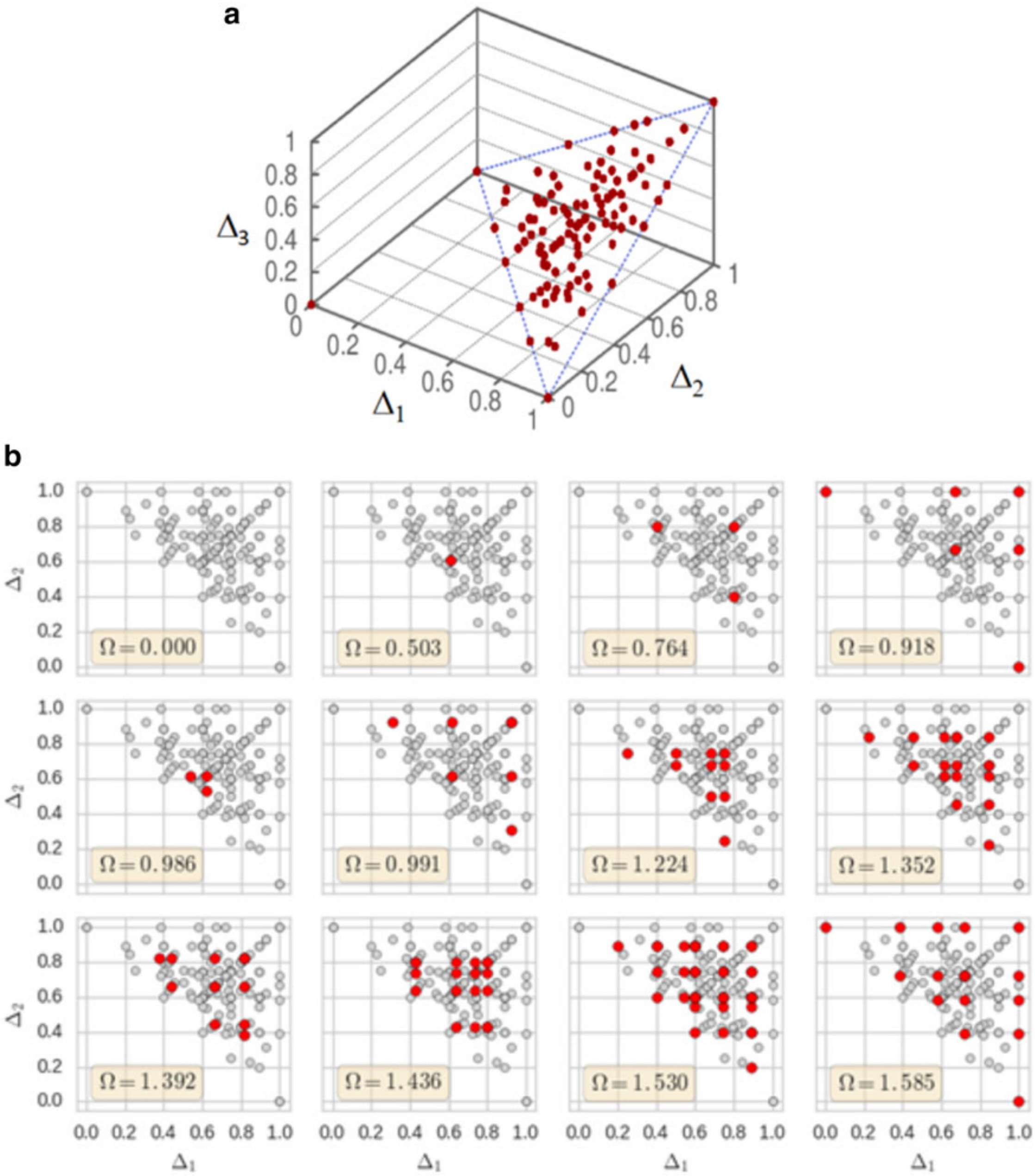


FIG. 3. Function classes (3×3) on the landscape. Each spot in both panels represents a function class, or family. **(a)** The information landscape shows the orientation of the plane with respect to the 3D landscape. **(b)** A set of 12 panels, one each for the complete set of possible values of the multi-information, Ω, for the 3×3 functions. The plane is the projected diagonal plane of the 3D landscape, the gray spots are the same for each panel and show the positions of all of the families of functions. The red spots are the families specific for each specific value of Ω. The upper left panel has no function as the information content of the uniform functions is zero, and all Δ's are zero. 3D, three-dimensional.

to emphasize the composite and subtle nature of the several factors that determine what ''noise'' is. This quantity can therefore only be inferred from the data when we can explicitly define the character and degree of the dependencies we are including. If we only consider direct pairwise effects, for example, from each of two loci on a phenotype, then the interaction between these loci affecting a phenotype (what we call a three-way dependency) as well as any other more complex interactions will contribute to the ''noise.'' Likewise, if we only include the effects of the genetic variants that we have ascertained, the loci not included will contribute to the ''noise.'' Any variables or effects not included can potentially contribute to

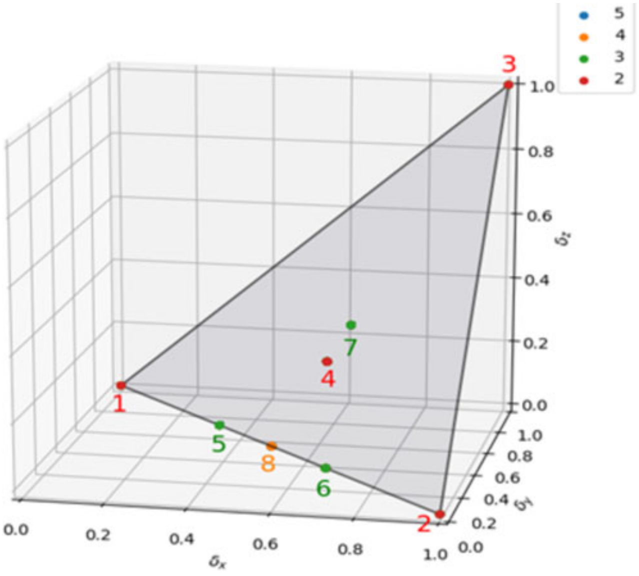


FIG. 4. The information plane for haploid genetics, binary genetic variables. The color-coded points show the locations of the function families corresponding to the alphabet size of the phenotype, as indicated in the legend panel at the top right. Families 1–4 correspond to a binary phenotype alphabet. Families 5–7 are added for a three-letter alphabet, and family 8 is added for a four-letter alphabet. The blue dot in the legend is not seen since it does not correspond to any specific family. The five-letter alphabet functions all fall into the previous eight families. While the limit is eight families, as the alphabet size increases the number of functions in every family grows. The families 1, 2, 5, 6, and 8 are functions with only pairwise interactions ($\delta_Z = 0$).

the ‘‘noise.’’ All unknown genetic variants and all other unknown environmental factors may contribute to what we call the ‘‘noise’’ in this point of view. In this way we can both more tightly define the quantitative nature of genetic penetrance and also provide a well-defined method for a data-driven estimate of the key quantities. We therefore have two fundamental steps in a general method for the inference of the relevant dependencies: first, the detection of levels of dependencies using the information theory measures, followed by the inference of the functional nature of these dependencies and the ‘‘noise’’ level. The ‘‘noise’’ plays a critical role in determining the penetrance of genetic effects.

5.2. *Probabilistic model*

The discrete functions of three variables, interpreted as distributions, are illustrated in the above landscapes (Figs. 3 and 4), where the information measures are calculated from these functions. Since any phenotype is not fully determined by genetic functions, ‘‘noise’’ is recognized as an important factor in quantitative genetics, as we emphasized above. What we mean specifically by noise, however, includes unknown sources of effect, as well as truly stochastic factors, both biological and technical. The mathematical noise function we use here, ε, thus arises from several sources, specifically including the following six:

1. Measurement errors in any of the variables, both phenotypes and genotypes.
2. Environmental influences on the phenotypes.
3. Epigenetic effects.
4. Stochastic developmental and physiological effects.
5. The effects of uncharacterized genetic variants (rare Single-Nucleotide Polymorphisms [SNPs], Copy Number Variations [CNVs], etc.—anything not included in our genetic variables), interactions that involve more than two genetic loci, which we do not include here, and weak effects from other two- and three-variable tuples that are below a statistical threshold for consideration. These are all what are usually called genetic background effects.
6. Sampling noise (allele frequencies vs. subjects, etc.), purely statistical fluctuations.

The ‘‘noise’’ as defined here, of course, is actually not noise in the usual sense of the word, but is the composite of all unknown influences as well as truly stochastic inputs.

The discrete functions represent a vanishingly small fraction of all possible information functions. However, they can be used effectively to describe real genetic effects, and generalized by adding a noise function that modifies the probability of occurrence of each possible alphabet value of the phenotype. This allows us to flexibly represent general distributions for any specific alphabet size, and thereby defines the ‘‘noise’’ in our functions as described above. It is clear that the locations of these general functions on the information landscape are continuously distributed, as illustrated in Sakhanenko et al. (2017) where we introduced random noise into the discrete functions. Here we introduce a systematic formulation that combines discrete functions with noise.

In general terms, the relation between genetic loci and phenotypes is embodied in this formalism in discrete valued *loci/phenotype arrays*: $f(X_1,\ X_2,\ \ldots X_n;\ \Phi)$, where $\{X_i\}$ is the set of n genetic loci and Φ is the phenotype. For two genetic loci, this is identical to a ''gene/phenotype table.'' When we take account of ''noise'' distributions, we add a uniform distribution to these arrays and form the genotype/phenotype arrays. The relation between these for n variables is simply

$$G(X_1,\ X_2,\ \ldots X_n;\ \Phi,\ p) = pf(X_1,\ X_2, X_3 \ldots X_n;\ \Phi) + (1-p)\varepsilon \tag{17}$$

where ε is a uniform array representing uniform random noise, and (1-p) is the ''noise'' level. The parameter p is the *penetrance*.

Since in the three-letter alphabet the discrete functions determine a third variable as a function of two others, the functions can be represented by $3\times3\times3$ arrays, G_{ijk}. This array can be understood as the probability mass function for the genetic variables specified by i and j, and the phenotype variable specified by k. This is composed to two other $3\times3\times3$ arrays: the function array f_{ijk}, which is nonzero only when $f(X_1,\ X_2) = k$, and the noise array ε.

For the genotype/phenotype array, the fractional balance between the ''noise'' and the discrete function is a variable factor we call penetrance, in keeping with the usual use of the term in genetics. If the penetrance is 1, there is no confounding noise, and if it is small, the genetic function plays that correspondingly small role in determining the phenotype. Note that if the penetrance is small, the significance of the genetic effect is also small. Thus, there is a clear relation between penetrance and the p-value of the effect. This relation will be explicated further elsewhere.

We assume here that a full penetrance effect can be described by a single discrete function. It is possible that some linear combinations of discrete functions could be useful in some cases. We do not consider this more complex extension further in this article.

It is important to see what happens to the coordinates as the penetrance decreases (''noise'' increases). To see what the delta coordinates are in the information landscape for functions with low penetrance, we examine the limiting ratios of the information functions. Since we cannot calculate deltas for the uniform distribution, consider distributions infinitesimally close to the uniform distribution or close to zero penetrance.

$$\left[\mathrm{P}^{\mathrm{rand}}_{\mathrm{ijk}}\right] = \frac{1}{27}\left[\begin{bmatrix} 1+\epsilon & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},\ \begin{bmatrix} 1-\epsilon & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\right], \tag{18}$$

and calculate the delta coordinates in the limit $\epsilon \to 0$. The above array can be used to calculate the corresponding joint entropy, for example, for the 3×3 case.

$$H_{123} = -\frac{25}{27}\log\frac{25}{27} - \left(\frac{1+\epsilon}{27}\right)\log\left(\frac{1+\epsilon}{27}\right) - \left(\frac{1-\epsilon}{27}\right)\log\left(\frac{1-\epsilon}{27}\right)$$

Note that we could change Equation 18 to add/subtract ϵ from different elements along a different dimension, and that this would lead to the same value for H_{123} and other constituent entropies. The delta-coordinates can be calculated from these entropies. The first delta-coordinate is

$$\delta_1 = \frac{-H_1 + H_{12} + H_{13} - H_{123}}{\sum_i H_i - H_{123}}$$

However, both the numerator and denominator of this expression go to zero in the limit when $\epsilon \to 0$, so we take the limit using L'Hospital's rule. The first derivatives are also each zero, but using the second derivatives yields the limit:

$$\lim_{\epsilon\to0}\delta_1 = \frac{8/81}{32/243} = \frac{3}{4}$$

Note that this analysis does not strictly prove that this limit is the same for all possible forms of structured noise, but serves as an analytical verification that agrees with all previously observed numerical results (Sakhanenko et al., 2017). This location $\left(\frac{3}{4}, \frac{3}{4}, \frac{3}{4}\right)$, on the information plane, has very particular properties that need to be carefully considered. When the array is completely dominated by the uniform probability, the ''noise'' completely swamps out the information content of the functions, and the genetic information has no effect on the phenotype. This point corresponds to a value of the penetrance, p, of zero. It is the location on the information plane that we called the ''black hole'' previously (Sakhanenko et al., 2017). This suggests that as noise increases, the functions all move their positions on the landscape, and they eventually converge on this spot.

5.3. An algorithm for inferring genotype/phenotype arrays

Since the relation between genetic loci and phenotypes is described by discrete valued *loci/phenotype arrays* (see Equation 17), once we have used the information measures to determine that there is a significant dependence for a given set of variables, we need to infer the function itself to understand what the data imply. However, since the array is not described by a discrete function alone, we also need to infer the level of the essential ''noise'' distribution. As described above, together these components, the discrete function and the ''noise'' level, described by the penetrance, form the genotype/phenotype arrays, where the function ε is a uniform random ''noise'' function, and (1-p) is the noise level, and the parameter p is the penetrance (Equation 17). We will henceforth write the arrays using indices that range over the variables and the alphabets. Thus $f(X_1,\ X_2,\ X_3\ \ldots X_n, k)$ is written as $f_{i_1 i_2 i_3 \ldots i_n\, k}$, where the $\{i_1\}$ are genetic variant indices, and k is an alphabet index.

$$G(X_1,\ X_2,\ \ldots X_n;\ p) = pf_{i_1 i_2 i_3 \ldots i_n k} + (1-p)\varepsilon \tag{19}$$

Given a data set and a significant tuple of variables (the dependence to be analyzed), there is a simple way to infer the function and p. The problem can be posed as follows for a three-way dependence. Let us represent the data by the data frequency array for a phenotype as a function of two genetic variants

$$D_{ijk} = \left\{ \begin{bmatrix} d_{111} & d_{121} & d_{131} \\ d_{211} & d_{221} & d_{231} \\ d_{311} & d_{321} & d_{331} \end{bmatrix}, \begin{bmatrix} d_{112} & d_{122} & d_{132} \\ d_{212} & d_{222} & d_{232} \\ d_{312} & d_{322} & d_{332} \end{bmatrix}, \begin{bmatrix} d_{113} & d_{123} & d_{133} \\ d_{213} & d_{223} & d_{233} \\ d_{313} & d_{323} & d_{333} \end{bmatrix} \right\} \tag{20}$$

where all variables range over three-letter alphabets. This frequency array is defined to be normalized so that the sum of all components is 1.

Let us assume for the moment for simplicity of explication that the allele frequencies are equal. We will modify the resulting simple algorithm for nonequal allele frequencies later (note that this is moot for the haploid case we analyzed in the last section). A simple greedy algorithm for finding the most likely function, f_{ijk}, from the data d_{ijk}, simply identifies the maximum d_{ijk} for each letter of the alphabet k, and assigns a probability of one to that k and zeros to the other two for all i and j:

$$\forall i, j\ :\ f_{ijk} = \begin{cases} 1, & \textit{if}\ d_{ijk} = \max\limits_{t}\left(d_{ijt}\right) \\ 0, & \textit{otherwise} \end{cases}$$

The algorithm is ''greedy'' in the sense that it takes the largest value of d_{ijk} for each k and gives it a value of 1. This prescription is incomplete, however, in that the maximum within each value of k matrix may not be unique. In this case, we can choose the element to assign $f_{ijk} = 1$ randomly among the multiple maxima. We have not explored the quantitative impact of this source of noise but have found that for the large data sets explored so far, for example, the yeast data set in section 6, there are unique maxima. This is another case where the more samples in a population the more frequently the noise is suppressed. The estimate of p in Equation 19 is the average frequency of the array elements not assigned a value of 1 in f_{ijk}. If the expectation is taken over all array elements, since there are nine nonzero entries for f_{ijk}, we can write the expression for the penetrance, p, as

$$p \cong 1 - \frac{3}{2}\sum_{ijk}\left(d_{ijk}\left(1 - f_{ijk}\right)\right) \tag{21}$$

The algorithm yields the genotype/phenotype array

$$G_{ijk} = pf_{ijk} + (1-p)\varepsilon$$

There are many ways of characterizing the resulting fit—measuring how well the data are described by such an inferred function. In the spirit of the current formalism, we can calculate the Kullback/Leibler divergence between D_{ijk} and G_{ijk}, but a chi-squared test also works. Note that these arrays are normalized so that they can be treated directly as distributions.

5.4. The effect of allele frequencies

In the previous section, we made the simplifying assumption that the allele frequencies were equal to more clearly explain the process. The frequencies are, of course, hardly ever equal. To deal with this issue, we can make a simple linear transformation of the matrices of the array to account for unequal allele frequencies, which slightly complicates the algorithm, but is not a fundamental difference. More importantly, it is essential to note that allele frequency differences can have strong effects on all of the information measures. Among other difficulties, strong interactions between loci, three-way effects, can potentially be masked by the rarity of key alleles in these loci. There is little that can be done to avoid this problem if it fully masks the interaction signal, however, the detection of weak interactions should therefore be looked at carefully in the population to ascertain whether the allele frequencies are involved in determining the strength of the signal. Additional cautions can include segregating the sample population to focus on more genetically homogenous subpopulations. This can only help, of course, if the numbers are large enough to properly analyze a subpopulation, but it may be a necessary step to avoid missing significant effects.

5.5. A simplification: Transformations of three functions into two functions

Each of the discrete functions of two variables, $Z(X,Y)$, can be specifically transformed into a pairwise function without loss of dependency information. By this we mean expressing it as a function of a single variable that maps to values of the pair X and Y. For example, the 3×3 function we call an XNOR-like function (where XNOR is the logical complement of XOR) is represented by this matrix that defines Z values (the columns are X, and the rows are Y).

$$\begin{bmatrix} 2 & 0 & 1 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

A transformation that maps pairs of (X,Y) values into another variable, call it W, can represent the essential information in the function: $\left\{ \left(X_i, Y_j\right) \right\} \Rightarrow W$. What this means is that the matrix can be fully reconstructed by the mapping

$$\{(1,0),(2,1),(0,2)\} \Rightarrow 0$$
$$\{(2,0),(0,1),(1,2)\} \Rightarrow 1$$
$$\{(0,0),(1,1),(2,2)\} \Rightarrow 2$$

since this transformation simply yields values of the variable W such that $W=Z$. Every function can be mapped similarly. The resulting two-variable functions have the properties that the mutual information between W and Z is maximal. These transforms take these three-way functions into a space of pairwise functions with only pairwise dependencies. An important question, however, is whether the mapping of the three-variable function space into a pairwise function space of the same size exhibits symmetries and redundancies that reduce the complexity of the one-to-one transformations. In other words, are there a set of ''basis'' transforms that can distinguish each of the three functions from one another when mapped into the pairwise function space? The XOR and XNOR functions for three variables have no pairwise dependence at all and point out their importance for data analysis. In Appendix B we show in the three-variable case that the symmetric delta is maximized by this family of functions.

The information measures described here are sensitive to dependencies, but do not define the functional nature of the dependencies, the functions themselves. In other words, the mapping of functions into delta coordinates is many to one, as is evident from the multiple functions in a family of functions with the same coordinates. The geometry of the information landscape, however, can usefully limit the possibilities since the delta values define a location in the landscape and thereby restrict the functions that could be generating the observed dependence. The detection of dependence, localization on the information landscape, followed by the identification of the actual function that then leads to transformations from three-function space to two-function space are the steps in the process of complete characterization of genetic phenomena with two-loci functions affecting phenotypes. The next question then is, how can this paradigm best be implemented?

5.6. Inferring the genotype/phenotype function and penetrance in simulated data

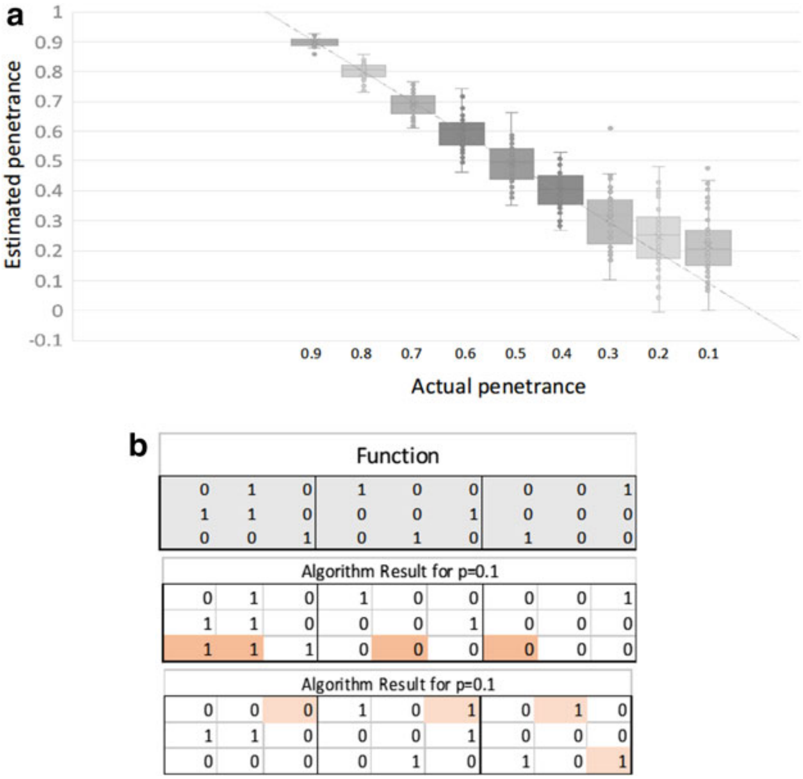
To illustrate the effectiveness of the simple algorithm described in the previous section, we created a simulated data set. We generated simulated data for 100 subjects. As test case, we used a specific discrete function and added a uniform ''noise'' function. The discrete function of three variables chosen in the case described here is shown in Figure 5. This function exhibits both pairwise and three-way dependence.

We generated the genotypes randomly, making the simplifying assumption of equal allele frequencies, and used this function to determine the phenotypes, then added uniform noise to the array using the relation of Equation 19 to generate the data set for specific values for the penetrance. Correction for allele frequencies is a simple linear transformation. The algorithm was used to infer the discrete function and estimate the penetrance. The results, both for this function and others not shown here, demonstrated that the algorithm works well to infer the exactly correct discrete function for all values of the penetrance greater than about 0.24. Penetrance levels less than this value (high ''noise'' levels) lead to some incorrect entries, as shown in Figure 5.

It is clear from these results that the simple algorithm provides a reasonably robust method for inferring a complex discrete function from data as well as estimating the penetrance. For larger data sets, of course, the thresholds for inference errors will be smaller than seen here.

5.7. Genetic heritability

The quantitation of heritability has been an important long-standing problem in quantitative genetics originally approached by consideration of variance. The ideas of what are called broad and narrow sense



Function

0	1	0	1	0	0	0	0	1
1	1	0	0	0	1	0	0	0
0	0	1	0	1	0	1	0	0

Algorithm Result for p=0.1

0	1	0	1	0	0	0	0	1
1	1	0	0	0	1	0	0	0
1	1	1	0	0	0	0	0	0

Algorithm Result for p=0.1

0	0	0	1	0	1	0	1	0
1	1	0	0	0	1	0	0	0
0	0	0	0	1	0	1	0	1

FIG. 5. Analysis of simulated data. **(a)** These are values of the penetrance calculated for 50 simulated data sets each for 9 values of penetrance, p. For all of these values, except the two right-most ($p=0.2$ and $p=0.1$), the greedy algorithm returned the exact function. **(b)** For these two there were a few errors in the function (top panel is correct function), as shown in these examples for two cases of $p=0.1$ simulations (the errors are highlighted).

heritability have their roots in the Fisher paradigm (Fisher, 1918; Wright, 1926). In the classical model, the components of the trait value (phenotype) are commonly these: the population mean, the genetic effect, and the ''error term.'' The assumption of normally distributed components, with no covariance, imposes a model dependence on the analysis. This model is convenient in that it implies that the variances add, but its validity is often in question. The broad sense heritability is simply defined as the ratio of the genetic variance to the population average, or the fraction of trait variance that is due to all genetics. In this connection, it is important to remind ourselves that Huang and Mackay have clearly shown that contributions of the additive, epistatic, and dominance variance components in the classical descriptions do not contribute only to these three respective variance components (Huang and Mackay, 2016). This means then that the variance components cannot be attributed to these model components, which has often been done in genetic analyses (Huang and Mackay, 2016). Genomic heritability is the fraction of the genetic variance that can be explained by regression on the markers and will only be quantitatively accurate when the genotypes of all causal variants are known. See Lstiburek et al. (2018) for a useful further discussion of heritability for panmictic populations.

When there is a way to determine the additive component of the genetic variance from separate experimental data, the fraction of the trait variance that is attributed to the additive variance is the classically defined ''narrow sense'' heritability. Classical methods often use the analysis of variance of full and half-sibling families and use maximum-likelihood methods for relatives with different degrees of relatedness to estimate this quantity.

In the information theory formalism, heritability, in the ''broad sense,'' can be reduced to a quantity that is actually rather simple to state. Since the total dependence, including all components additive and nonadditive alike, between a set of loci and a phenotype can be quantitated by the multi-information, we can use this quantity effectively to define heritability if we have calculated the penetrance for each set of loci.

It is important to emphasize that the total dependence, measured by the multi-information, includes all effects, both additive and nonadditive effects. For a given phenotype then, we propose to define the heritability as the ratio of the total of all the dependencies (this means the multi-information for all subsets of dependent variables affecting the phenotype), times the respective penetrance for each subset, divided by the ''maximum possible'' dependence for the same sets of loci. The maximum possible dependence is clearly the multi-information assuming full penetrance for all dependencies, and therefore, no effective ''noise.'' Therefore, the heritability, $\mathcal{H}$, can be written by the expression

$$\mathcal{H}_\phi = \frac{\sum_\tau p_\tau \Omega_\tau(\phi)}{\sum_\tau \Omega_\tau(\phi)} \tag{22a}$$

where τ is a subset of variables containing the phenotype variable and genetic loci. These are the sums over all subsets of the full set of all genetic loci exhibiting dependence for the specific phenotype, ϕ. In the numerator with the corresponding penetrances, and in the denominator without. If the penetrance is full for all determinants of the phenotype, the heritability reduces to one full heritability. This means that there are no environmental effects, unaccounted for loci or subsets of variants, or other sources of ''noise.'' In order for this to be a valid heritability for trait ϕ, of course, all possible genetic variable subsets must be included for which $\Omega_\tau(\phi)$ is nonzero. Equation 22a is then a valid and rigorous abstraction, but one that requires all possible multilocus effects to be quantitated to actually calculate.

One might think that since the sum over dependent tuples may not be disjoint, having some overlaps, the dependency could be overcounted. In other words, some loci may participate in several subsets of dependent loci. This kind of potential overcounting is not a problem, however, since the measures are weighted by the penetrance and normalized by the total sum, including all possible overlaps. Note that in this definition of heredity we need not assume linkage equilibrium. In fact, Equation 22a applies in the completely general case, with or without LD.

Practically we must limit the sum to those subsets of variables whose dependencies are detectable and significant, so the criteria for significance must also enter the determination of heritability. This is not because weak dependencies do not count, but because the calculation of p_τ can only be accurate if the dependence is significant. This definition is different from the classical description then in yet another way. We are calculating the heritability of traits based on all the variants considered in the analysis, while the variance form purports to include all genetic effects but is dependent on the unknown range of genetic differences in the population considered.

If the dependencies for a phenotype were all single-locus effects (pairwise dependence), then the heritability would only be a function of the mutual informations between these loci and the phenotype, ϕ:

$$\mathcal{H}_\phi(single\ locus) = \frac{\sum_i p_i I(i, \phi)}{\sum_i I(i, \phi)} \tag{22b}$$

where $I(i, \phi)$ is the mutual information between the loci and the phenotype ϕ.

In this article, we also restrict the sum to triplets, subsets of three variables, two loci, and the phenotype of interest, although Equation 22a is certainly valid for any size of subset τ, and any number of genetic loci. Since the composition of dependence for each triplet can be clearly separated into the components due to single-locus and two-loci dependence, as long as the two loci are independently segregating ($I(i, j) = 0$) we can also then separate the heritability into two components by separating the sum in the numerator into two parts, the pairwise or additive effect and the three-way effect. In this case, from Equation 9b it is clear that for the heritability limited to two-locus epistatic interactions (no single-locus additive effects), $\mathcal{H}_\phi(triple)$, we have

$$\mathcal{H}_\phi(triple) = \frac{\sum_\tau p_\tau \Delta_3(\tau)}{\sum_\tau \Omega_\tau} \tag{23}$$

where τ indicates all triple dependencies. This formulation provides a rigorous and complete description of heritability given the division between the genetic determinants and the unknowns, the ''noise.'' It also provides a practical way to calculate the heritability under specific assumptions. Contrast this with Fisher's heritability, *in the broad sense*, which is the ratio of the variances of phenotype to genotype in the population. Narrow sense heritability is more important in the sense that it quantitates the proportion of the phenotypic variation that is transmitted from parents to offspring (Lstiburek et al., 2018). The argument for this interpretation that ignores epistatic effects, which are frequently disrupted by segregation, is plausible, but it is certainly incomplete.

5.8. Protective alleles: interactions that nullify effects

The interaction of two loci, of course, means that each locus may modify the effect of the other in some way. In medical genetics, it is becoming increasingly clear that there are potentially severe effects of pathological variants that are not realized in phenotypes. This means that the genetic background of single or multiple loci is providing an effect that ''protects'' the subject from the pathology. This is an important area of research at the moment. These recent examples for Alzheimer's disease are emblematic of the approach (Ridge et al., 2017; Arboleda-Velasquez et al., 2019), which promises to provide biological insights into the mechanisms of pathology, and therefore, the nature of the genetic functions expected to be encountered in the analysis of genetic data is worth investigating. What is clear from our formulation is that certain discrete functions involving two loci can exhibit a protective-like character, which can be characterized.

To make this more precise, and to illustrate this kind of interaction in our formalism, we look at a specific concrete case to examine the instance of protective alleles. The hallmark of protective effects is easily described in terms of the genotype/phenotype table. For simplicity let us consider a binary phenotype where 1 is a negative phenotype, a disease state, and 0 is normal. The effect of a protective allele then simply means that one variant of gene A has the effect of reversing the disease effect of gene B and making the phenotype normal. This can be viewed as a kind of dominance, but a simple model example illustrates the point. A model that shows a protective effect given the gene assignment above is illustrated in Table 3.

Table 3. A Genotype/Phenotype Table (100% Penetrance) Illustrating a Protective Effect of Gene A Alleles on the Disease-Causing Alleles of Gene B

$B\setminus$A	0	1	2
0	0	0	0
1	1	0	0
2	1	1	0

Phenotype 1 indicates diseased, 0 is normal.

The minor allele of gene B (one or two copies) is assumed to cause the pathology except that it is neutralized in the presence of the minor allele of gene A (one or two copies), which is the protective allele. There are of course other functions that exhibit such effects.

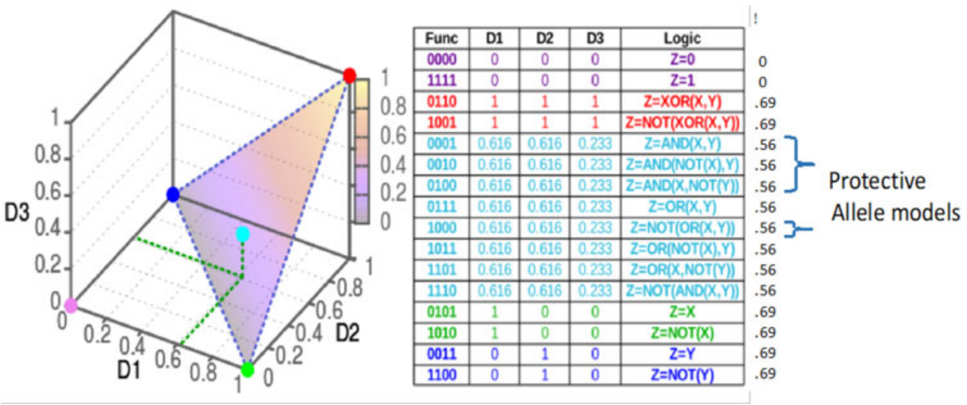
To illustrate the systematic effect in a very simple case, we consider the haploid genetic case with binary phenotypes, where there are only 16 possible genetic models (Fig. 6). Only 4 of the 16 possible 2×2 genetic models exhibit protective effects.

6. ANALYZING A YEAST GENETIC DATA SET

To illustrate the application of the information theory approach to quantitative genetics, we analyze a data set of haploid data from a large yeast cross generated by Kruglyak and colleagues (Bloom et al., 2015). The data consist of 4390 haploid strains resulting from the cross of a wild vineyard strain, and a widely used laboratory strain of *S. cerevisiae*. This is an F2 cross, so that the recombinations between the two parental chromosomes occur in a single meiosis event for each of the resulting strains. The resulting haploid strains are essentially the gametes from the hybrid F1 strains. The data include genotypes of all 4390 strains, at 28,820 SNP positions, and 20 phenotypes, average growth rates under different conditions and in the presence of different compounds. We have restricted our use of the data to those phenotypes that showed a relatively high reproducibility in replicates. We used only those phenotypes whose replicates exhibited highly consistent correlation coefficients. These criteria, a replicate correlation coefficient above 0.8, selected 4 of the 20 phenotypes reported by Bloom et al. (2015). We report the analysis of two of these four phenotypes: growth in the presence of neomycin (correlation coefficient 0.86) and copper sulfate (0.82).

6.1. Genetic dependencies

We calculated the pairwise effects, mutual information, between single genetic variants and the phenotype, and the measure of three-way effects, using a representative set of 100 variant markers across the genome. To calculate the three-way interactions accurately, we wanted independently segregating markers, so we selected a set of 100 markers that were isolated by iteratively eliminating one of each pair of markers that had a mutual information of more than 0.05. The markers were widely spread, and we calculated the recombination frequencies between each pair of neighboring markers to assess the statistics of segregation. The results for pairwise and three-way genetic dependencies for the two phenotypes are shown in Figures 7–10. For our significance calculations, we follow the permutation strategy proposed by Churchill and Doerge (1994): we shuffle the input data, breaking the connections between genetic markers and phenotypes, compute the dependency scores of all shuffled tuples, and count how many randomized scores are above the original score of interest. We repeat this procedure 100 times tallying the number of scores



Func	D1	D2	D3	Logic	
0000	0	0	0	Z=0	0
1111	0	0	0	Z=1	0
0110	1	1	1	Z=XOR(X,Y)	.69
1001	1	1	1	Z=NOT(XOR(X,Y))	.69
0001	0.616	0.616	0.233	Z=AND(X,Y)	.56
0010	0.616	0.616	0.233	Z=AND(NOT(X),Y)	.56
0100	0.616	0.616	0.233	Z=AND(X,NOT(Y))	.56
0111	0.616	0.616	0.233	Z=OR(X,Y)	.56
1000	0.616	0.616	0.233	Z=NOT(OR(X,Y))	.56
1011	0.616	0.616	0.233	Z=OR(NOT(X),Y)	.56
1101	0.616	0.616	0.233	Z=OR(X,NOT(Y))	.56
1110	0.616	0.616	0.233	Z=NOT(AND(X,Y))	.56
0101	1	0	0	Z=X	.69
1010	1	0	0	Z=NOT(X)	.69
0011	0	1	0	Z=Y	.69
1100	0	1	0	Z=NOT(Y)	.69

FIG. 6. Four of the 16 2×2 genetic models show protective effects. The functions are shown in linear form and color coded according to the families as marked on the information plane.

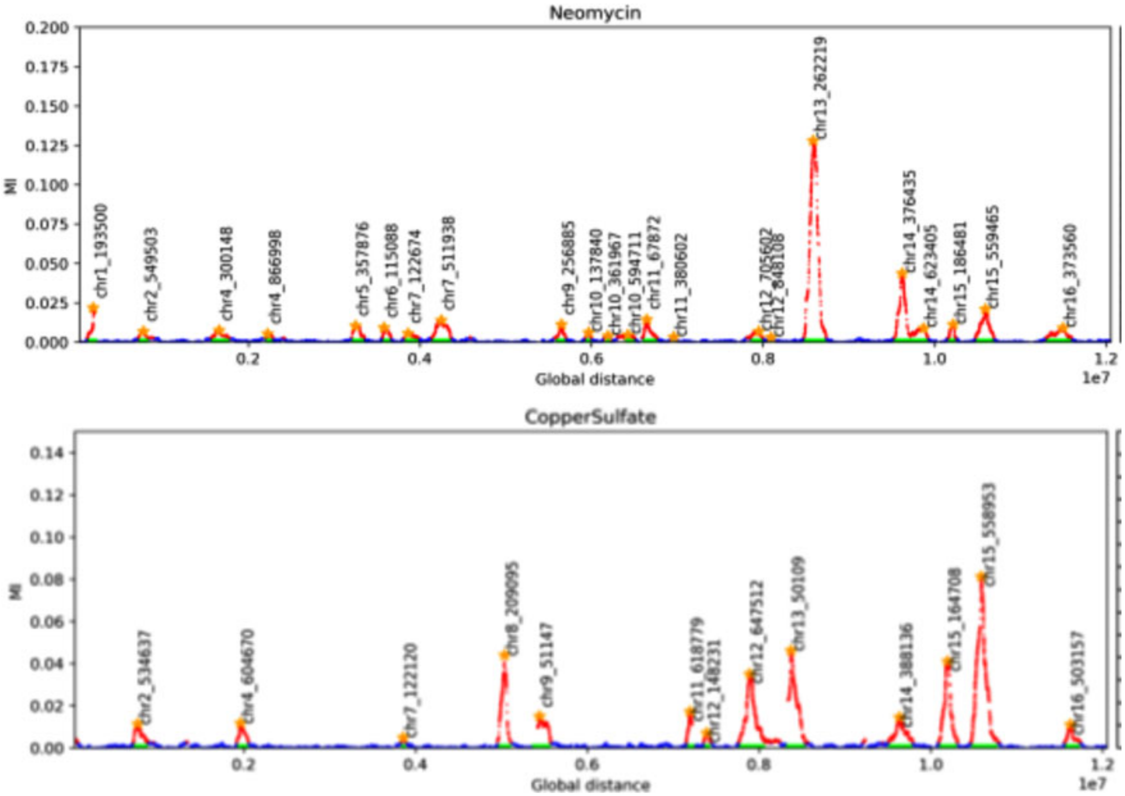


FIG. 7. The pairwise peaks for the genetic determinants of two phenotypes. The locations indicated are the chromosomal coordinates of the highest scoring marker in the peak.

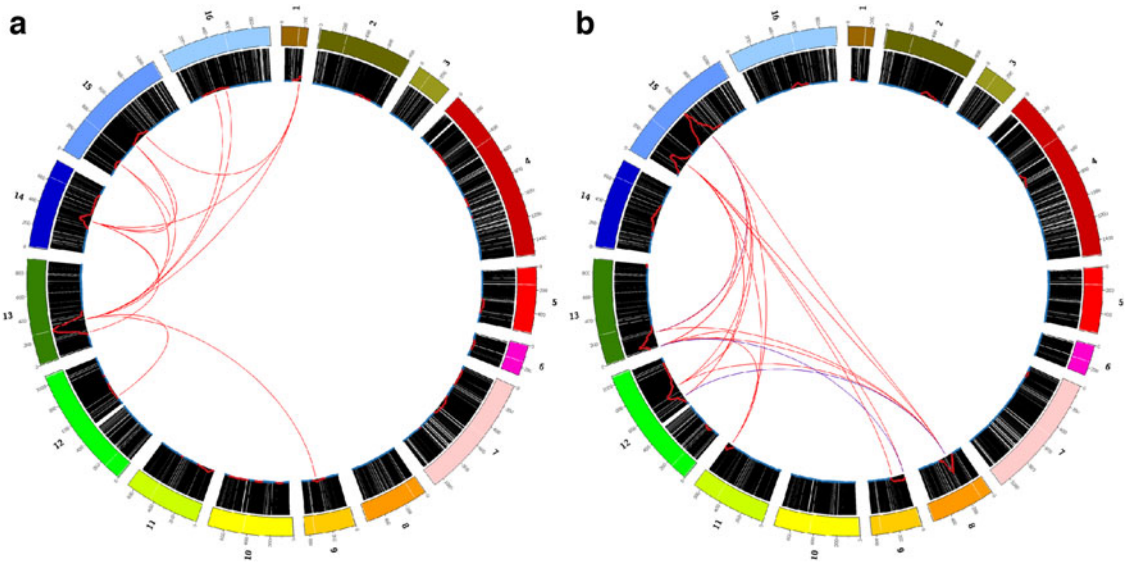


FIG. 8. Genetic effects for two growth phenotypes. The full genome is shown with the single- and two-locus effects. The pairwise peaks for these phenotypes, as shown in Figure 7, are indicated as the red curves in the black band (using all 28,820 markers.) The variant pair interaction effects for these phenotypes are indicated by the internal red lines (all interacting pairs are shown in the Appendix Tables D1 and D2) indicating the significant three-way dependencies between the two markers at the ends of the line and the phenotype, indicating genetic loci interacting. **(a)** Genetics of growth on neomycin and **(b)** growth on copper sulfate.

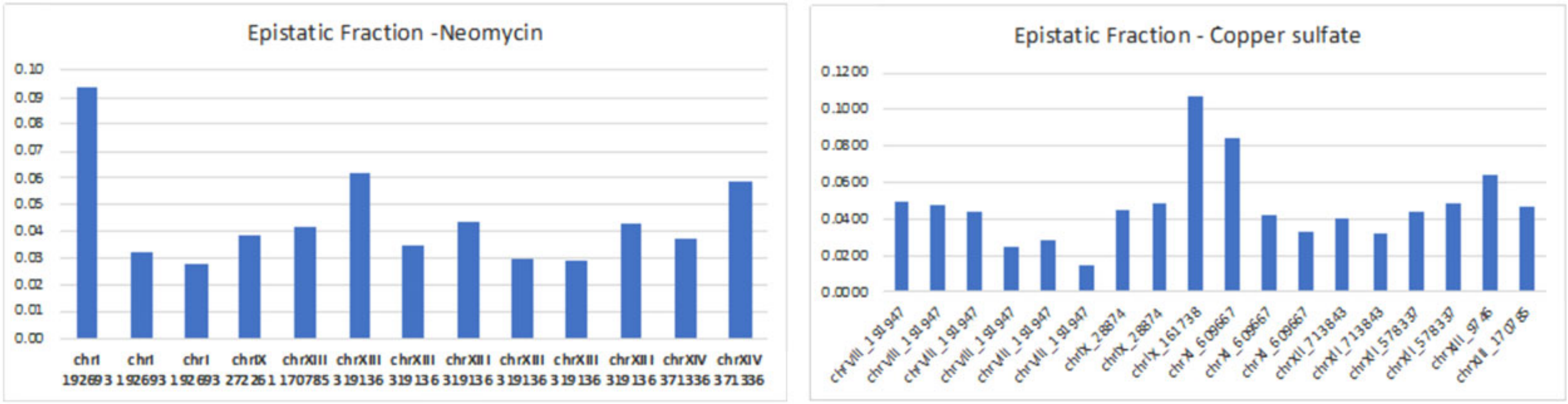


FIG. 9. Epistatic fractions compared. The bars represent the fractional epistatic effect for the interactions in the order listed in Appendix Tables D1 and D2. For brevity, the marker indicated below each bar is the one listed in the left-hand column of the tables, and represents the pair.

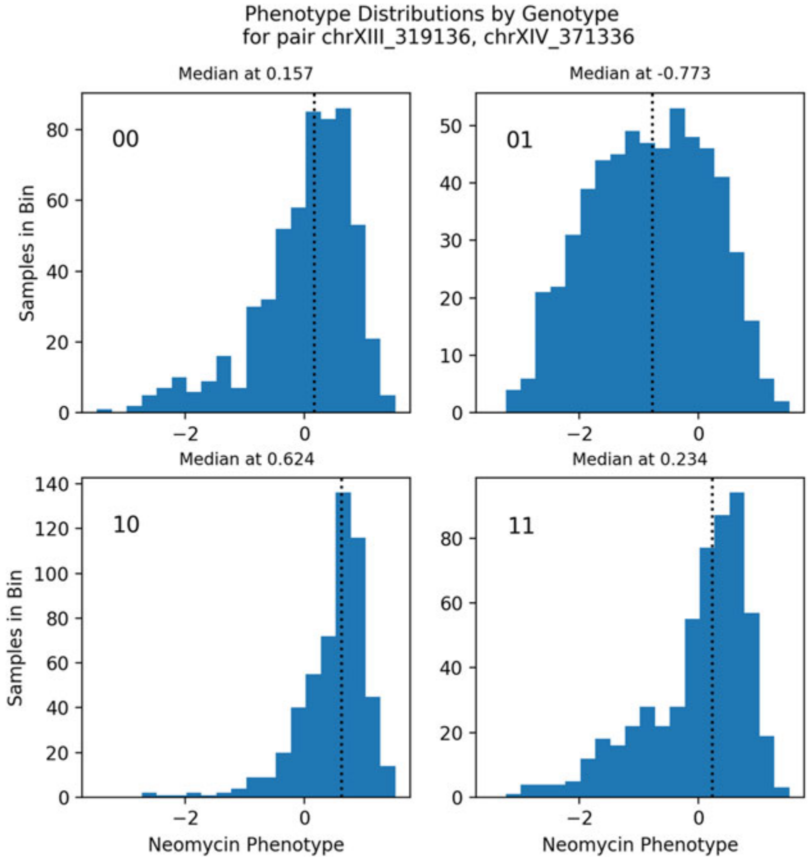


FIG. 10. Phenotype distributions by genotype. We examined the tuple with the highest value of 0 for the neomycin phenotype: loci chrXIII_319136 and chrXIV_371336. The panels show the phenotype distribution for each genotype (e.g., plot 01 shows samples with chrXIII_319136 = 0 and chrXIV_371336 = 1).

above the score of interest. The *p*-value is then the fraction the exceeding randomized scores take in the total number of tuples times 100.

The two panels in Figure 7 show the pairwise peaks resulting from plotting the mutual information for the entire set of 28,820 genetic markers for each of the two phenotypes. The entries in Table 4 show the location of the highpoint of each peak, the standard deviation, and the width. The width of the peak is defined by the outermost boundary of the peak determined by the locations of the last significant marker by mutual information on each side of the peak. The peak widths in this case are largely due to the cosegregation of contiguous blocks in the F2 meiosis.

In Appendix D, Appendix Tables D1 and D2, we indicate the quantitative results for the interchromosomal interactions shown in Figure 8. We decided to omit intrachromosomal interactions because in this cross with only one meiotic recombination, the correlated blocks of markers are significant and even the widely spaced markers used here have some residual correlation. The interchromosomal pairs have no such systematic correlations because of the nature of the cross, and thus, the calculations of their epistatic effects are most accurate, as explained above. The fractions of the interactions that are attributed to epistatic and additive effects are indicated in the final two columns of the Appendix Tables D1 and D2. There are variable levels of epistasis evident, but they are all below 10%.

A gene known to affect copper resistance, the well-known metallothionein gene, *CUP1*, which is in the peak on chromosome VIII, participates in several interactions. Note for all the interactions described here that since the markers used for this analysis have relatively low resolution, two neighboring markers may well indicate the same epistatic interaction. This is particularly likely when these markers are within the same mutual information peak, but we do not attempt to separate these effects here.

Table 4. Pairwise Peaks: Global Location, Standard Deviation, and Widths

Neomycin			*Copper sulfate*		
Peak	*STD*	*Width*	*Peak*	*STD*	*Width*
chr1_193500	22923.9	87228	chr2_534637	48233.6	169846
chr2_549503	32012.6	123030	chr4_604670	36277.6	140271
chr4_300148	55234.1	220149	chr7_122120	15259.7	52115
chr4_866998	22360.5	83612	chr8_209095	36374.4	159134
chr5_357876	38574.9	148433	chr9_51147	44126.4	151924
chr6_115088	33540.9	102487	chr11_618779	19211.5	69362
chr7_122674	41580.8	166285	chr12_148231	21742.2	83786
chr7_511938	60005.7	226469	chr12_647512	87081.0	309232
chr9_256885	31559.4	117476	chr13_50109	50420.5	203113
chr10_137840	18979.8	68372	chr14_388136	82244.5	286873
chr10_361967	18761.3	66082	chr15_164708	56742.5	212833
chr10_594711	30922.9	113917	chr15_558953	82886.9	311335
chr11_67872	48640.4	170876	chr16_503157	45362.8	178043
chr11__380602	1230.0	4248			
chr12_705602	39973.0	139637			
chr12_848108	2503.8	8462			
chr13_262219	80355.8	263928			
chr14_376435	47175.2	190403			
chr14_623405	48656.3	204340			
chr15_186481	24329.6	88475			
chr15_559465	60872.8	216317			
chr16_373560	87976.6	296871			

The coordinates here are the global coordinates in base pairs, beginning with chromosome I. The widths and the STD are in base pairs. The width is the distance between the two extreme, significantly scoring markers.

STD, standard deviations.

There are several notable features shown in the data of Figure 9, which compares the epistatic fractions for the two phenotypes. These quantities are the fractions of the total dependence of each of the two genetic loci that are attributed to their interaction. Note primarily that the epistatic fractions vary several-fold, and the variation among them includes interacting pairs with the same marker, so that the degree of epistasis is clearly determined by both markers. The left-most six interactions of the copper sulfate panel, for example, indicate the interactions with the marker on chromosome VIII near the *CUP1* locus.

To see an example of the detected epistatic interactions in the data, we analyzed the tuple that had the largest detected multi-information, Ω, with the neomycin phenotype: loci chrI_319136 and chrXIV_371336. As shown in Appendix Table D1, this tuple has a multi-information of $\Omega = 0.1266$, and an epistatic fraction of only 0.072 (i.e., an additive fraction of 0.928). The phenotype distribution for samples of each unique genotype is shown in Figure 10. Even with this large additive fraction, it is clear from the phenotype distributions that we cannot consider this interaction to be entirely additive. Genotype 10 has the highest median phenotype at 0.624. Flipping either locus to 00 or 11 results in medians of 0.157 and 0.234, respectively (decreases of 0.467 and 0.39). If these effects simply added, we would expect a decrease of about 0.86. The median effect on phenotype 01, however, is a decrease of nearly 1.4, well beyond what would be expected from an additive effect. This is corroborated by a qualitative assessment of each distribution: the distribution of genotype 01 is markedly different from that of any other genotype, implying that even the relatively small epistatic interactions that we detect have a real and noticeable consequence.

To verify the statistical significance of this epistatic effect, we performed a permutation test that shuffled the data while preserving pairwise relationships. Specifically, for each binned phenotype label, we randomly shuffled the genotype labels against each other; this preserves $p(x,z)$ and $p(y,z)$ while randomly permuting $p(x,y)$, and thus perfectly preserves any pairwise genotype/phenotype relationships within the

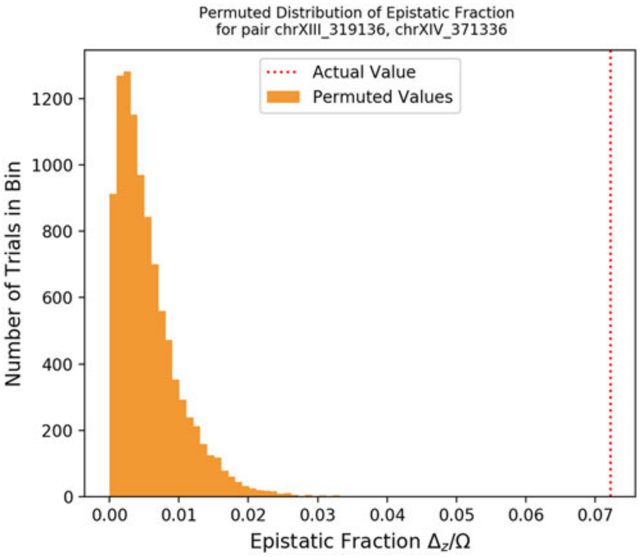


FIG. 11. Evaluating statistical significance of the epistatic effect for pair of loci chrXIII_319136 and chrXIV_371336. The distribution of 10,000 epistatic fractions calculated on permuted trials. The actual epistatic fraction is indicated by the dotted line.

data. This was repeated for 10,000 permuted trials, and the calculated distribution of epistatic fractions is shown in Figure 11. The actual epistatic fraction of 0.072 was substantially larger than any values obtained through these permuted trials, indicating that this epistatic effect is statistically significant.

6.2. Pleiotropic effects

Recall that pleiotropy is the effect in which a single genetic locus causes multiple phenotypic effects. For three-way dependencies, this means a triple set of variables showing a dependency that consists of a single genetic locus and two phenotypic variables. In the yeast data set, we are considering this is reflected in a pattern that is easy to observe. Consider Figure 8 and ask what a pleiotropic effect would imply. One instance would be, for example, a situation in which a single pairwise dependence is evident in both the neomycin- and the copper sulfate- resistant phenotypes. There are actually several such cases, but the most evident one that is used as an example are the two single-locus peaks that appear on chromosome XV in the figures showing both phenotypes, Figure 8a and b. The tables in Appendix D illustrate pleiotropy more specifically, identifying these two different variants that are coidentified for both phenotypes: these markers: chrXV_605280 and chrXV_189231. If we carry out the corresponding triple dependence measure using the symmetric delta, we do see these same loci showing up (results not shown). The symmetric delta is, of course, applicable for any combination of variables to detect collective dependence and the occurrence of pleiotropy, among two phenotypes and one genetic variant, is a collective dependence just as is the dependence among two genetic loci and one phenotype.

7. DISCUSSION

With the formulation presented here, we have begun to build a structure based on information theory for describing and analyzing data in quantitative genetics. The advantages of an information theory approach to quantitative genetics are multiple and include the following: a model-free agnostic approach to multivariable dependence, the simplicity of formalism, and the fundamental separation of dependence detection from the ascertainment of the functional nature of the dependency. There are important differences from the classical formulation: (1) GWAS includes a tacit assumption of linearity (the usual association studies largely make use of correlation coefficients, which by their nature quantify linear dependence, and thereby tacitly assume a linear model), and (2) variance component analysis assumes other tacit model assumptions discussed in the introduction.

These differences from the classical formulation permit us to formulate important genetic quantities in a direct, simple, and calculable way. This contrasts with the classical formulation, for example, with respect to the variances of phenotypes and genetic variance and the additivity of variances on the assumption of Gaussian distributions, and other more subtle assumptions pointed out by Huang and Mackay (2016). The assumption about which components contribute exclusively to which variances (additive, dominant, and epistatic) is based on models that are generally not valid, and this flaw renders the inference of genetic architecture from variance component analysis invalid (Huang and Mackay, 2016). The conclusion is summarized in this article as follows: ''The crux of the problem is the undesirable features of the classical model as well as the alternative parameterizations that there is not a one-to-one correspondence between gene action at underlying quantitative trait loci and the partitioning of variance components except under very specific and restrictive circumstances.'' Many researchers have pointed to the need for an alternative approach to the classical model (Nelson et al., 2013; Huang and Mackay, 2016) and our article begins the construction of such a formulation using information theory.

We fully recognize that the formulation in this article, limited to panmictic populations, provides only the first steps and that more development, which we plan for future publications and encourage others to address, is definitely needed. It has been well remarked that no natural population is panmictic, and that LD, relatedness, and population structure must also be described in the information theory formulation. These issues are amenable to an information formulation, while some of them can be somewhat more complex. They have been approached in the use of information theory for discussions of evolution and populations previously (Ting, 1962; Han, 1980), and we will treat these additional issues in a future publication.

Since the formulation presented here uses discrete functions extensively, it is important to understand the advantages and limitations of this underlying structure. While there are a finite number of discrete functions for any finite number of variables and alphabets, there are an infinite number of possible (discrete plus continuous) distributions for any finite number of variables and alphabets. For example, there are $3^9 = 19{,}683$ 3×3 discrete functions (three variables and three-letter alphabets). This obvious distinction is important, and it is clear that the function-to-distribution mapping is not one-to-one—there are vastly more distributions than discrete functions. Since we have assumed no LD for the most part in this article, the ''information landscape'' for two variants was confined to a plane. When the genetic variants are not independent variables, and linkage is present, the ''information landscape'' is no longer a plane and requires a more complex description.

It is clear that the addition of ''noise'' to the discrete functions (see Sakhanenko et al. [2017] and section 5.1) generates distributions around the discrete functions at varying distances in the information landscape. The point where the uniform noise distribution fully dominates and masks all information content, the ''black hole'' of information (Sakhanenko et al., 2017), represents the single distribution with uniform probabilities—it has maximal entropies. It may seem extraordinary, however, that this point has finite and distinct coordinates in the information landscape, as we describe in section 5.2. This points to the fact that the landscape includes functions with a wide range of penetrance and heritability, and that they are not at all distributed continuously on the landscape.

What is important here is to clearly define the distinctions and classifications that information theory can provide, and what these measures do and do not distinguish. This is an important specific question. One striking example is that the distribution of the ''black hole'' (Sakhanenko et al., 2017) and the discrete function that is XOR-like (with no nonzero pairwise measures) have distinct and unique coordinates in the information space for three variables, but as the alphabet size increases without bound (allowing more and more precise definitions of the variable values) they converge in the normalized space coordinates. It is a notable and related fact that the maximization of the symmetric delta leads to exclusively three-way-dependent functional relationships, as shown in Appendix B.

The application of this information-based formalism to genetics brings forward a number of interesting and important relations. The quantitation of the penetrance, for example, is simple and direct, but depends on a very clear division between what variables are being included to make the inferences and what is being ignored and therefore being considered to contribute to the ''noise.'' We characterize the nonincluded variables with the ''noise,'' even if they include the more complex genetic interaction effects that are not included. This is obvious and commonplace, but our formulation forces it to be explicit in all cases. Likewise, the information theory expression for heritability is both direct and intuitive, but also depends, as it must, on the precise assumptions made, and on the significance of the dependencies. Here we have ignored any variant interactions among more than two genetic loci, and effects that involve three or more loci can therefore contribute noise directly to the penetrance and heritability calculations. We are thus forced to be explicit about what is meant by genetic background effects, which also include any effects of variants not considered variables in the analysis.

Since the yeast two-strain cross involves only a single meiosis per strain to produce the collection analyzed, individual marker segregation within chromosomes is not a good assumption since blocks of markers will necessarily segregate together, while individual chromosome segregation, on the contrary, is generally an excellent assumption, and is supported by the data. In future work we will consider the more complex cases where the possibility of nonindependent segregation, and nonuniform population structures, is included, as in human data. The extension of the formalism to quantitate these effects and their implications will be an important part of the full description.

Bloom et al. (2015) argue that the additive effects are much greater than the epistatic effects on the quantitative traits they measured using the variance component method, estimating a 9% overall interaction effect. Our results agree with this qualitative conclusion, but we calculate the interactions specifically for each pair of interchromosomal markers and found that the levels of significant interactions among variants are highly variable, but in this general range. Huang and MacKay (2016) specifically point out that Bloom et al. used the variance component calculations improperly to make this estimate. It is therefore not surprising to find quantitative disagreement with our results, even if the general quantitative ranges are similar. Results in similar F2 crosses in mice also show small epistatic effects as we would expect (Tyler et al., 2016). These authors use logistic regression methods to look at both pleiotropic and epistatic effects of pairwise identified loci.

We argue that the ability to quantitate the pairwise, additive, and epistatic effects unambiguously is definitely a significant step forward and provides a distinctly different and practical alternative to the classical model. Comparison of our specific results with those of Bloom et al. shows close agreement for the pairwise effects, but unfortunately there is no way to make the full comparison for multiloci effects for the overall estimate of additivity. The major difference with Bloom et al. (2015) is that we accurately calculate the specific fraction of additive and epistatic effects for each pair of loci.

The present contribution, a first step toward a full information theory of quantitative genetics, leaves a number of important problems unaddressed. These include the characterization of population structure, relatedness, and LD, as we have mentioned. Their incorporation into the formalism, so they are accounted for in the genetic inferences, will be an important improvement. Genes with significant linkage may also interact. This is often a problem in genetic analysis, the disentangling of interaction from disequilibrium, and the information theory formalism method needs to be extended to treat these cases accurately. While larger data set sizes are required to accurately assess more interacting loci than two, the formalism can certainly address these more complex dependencies, which will be more and more important in future as data sets of genetic information continue to grow rapidly. Future work will focus on addressing all these issues and incorporating extensions into a formalism that will provide a broadly applicable set of descriptive and analytic tools.

8. APPENDICES

8.1. Appendix A. Two propositions regarding key inequalities

Proposition 1. $\Omega \geq I_{13}, \Omega \geq I_{12}, \textit{and } \Omega \geq I_{23}$ (A1)

Proof:

We first prove that $\Omega \geq I_{13}$.

$$\Omega = H_1 + H_2 + H_3 - H_{123}$$
$$I_{13} = H_1 + H_3 - H_{13}$$

Subtract the lower from the upper, to get $\Omega - I_{13} = H_2 - H_{123} + H_{13}$. Since $\Omega \geq 0$ and $H_i \geq 0$, and since the sum of the entropies of any subset of variables is greater than or equal to the joint entropy, it is clear that $H_2 + H_{13} \geq H_{123}$, and the proposition is proven. Starting with I_{23} or I_{12} yields the parallel inequalities. There are, however, tighter bounds.

Proposition 2: If Ω_{123} is the multi-information for three variables, then

$$\Omega_{123} \geq I_{13} + I_{23}, \;\; \Omega_{123} \geq I_{12} + I_{23}, \; \Omega_{123} \geq I_{13} + I_{12} \tag{A2}$$

By definition $\Omega = H_1 + H_2 + H_3 - H_{123}$ and the mutual informations, and interaction information, I_{123}, are

$$I_{123} = H_1 + H_2 + H_3 - H_{12} - H_{13} - H_{23} + H_{123}$$

$$I_{12} = H_1 + H_2 - H_{12}, \text{ etc} \ldots.$$

So we have

$$\Omega = I_{12} + I_{13} + I_{23} - I_{123}$$

Since the recursion relations for the interaction information are

$$I_{12} - I_{123} = I_{12|3}$$

$$I_{13} - I_{123} = I_{13|2}$$

$$I_{23} - I_{123} = I_{23|1}$$

we can write

$$\Omega = I_{13} + I_{23} + I_{12|3}$$

$$\Omega = I_{12} + I_{23} + I_{13|2}$$

$$\Omega = I_{13} + I_{12} + I_{23|1}$$

Since by the definition of the mutual information $I_{12|3} = H_{1|3} + H_{2|3} - H_{12|3}$, and we have the identity $H_{2|3} + H_{12|3} = H_{1|23}$ by the basic probability identities. Since $I_{12|3} = H_{1|3} - H_{1|23}$ and $H_{1|3} \geq H_{1|23}$, $I_{12|3} \geq 0$, the above three equations imply,

$$\Omega \geq I_{13} + I_{23}$$

$$\Omega \geq I_{12} + I_{23}$$

$$\Omega \geq I_{13} + I_{12}$$

And the proposition is proved.

8.2. Appendix B. Maximizing the symmetric delta finds any XOR-like dependence first

Equation 9a implies that if the mutual informations are all zero or very small, for a triplet, the deltas must all be equal, or close to it. Furthermore, they must all be equal (or close to it for small mutual informations) to the multi-information, Ω. The product of the deltas, the symmetric delta, is maximized when the factors are equal given a constant sum. This means that maximizing the symmetric delta will tend toward picking out the functions with deltas that are closest, given that the total correlation or multi-informations, Ω, are constant. The XOR-like functions, which have no pairwise, but do have three-way dependence, therefore maximize the symmetric delta for given Ω. An easy calculation shows that the maximum of the symmetric delta, $\bar{\Delta}$, occurs when all factors are the same. If we define variable measures of the pairwise dependence, $\{x_i\}$ then

$$\bar{\Delta} = \Delta_1 \Delta_2 \Delta_3 = \Omega^3 \left(1 - \frac{x_1}{\Omega}\right)\left(1 - \frac{x_2}{\Omega}\right)\left(1 - \frac{x_3}{\Omega}\right) = \Omega^3 \prod_k \left(1 - \frac{x_k}{\Omega}\right) \tag{B1}$$

where $x_1 = \frac{I_{12}+I_{13}}{\Omega}$, $x_2 = \frac{I_{12}+I_{23}}{\Omega}$, $x_3 = \frac{I_{13}+I_{23}}{\Omega}$

For all triplets, indicated by k and l, we then can say that $\underset{k}{\bar{\Delta}} > \underset{l}{\bar{\Delta}}$ only if

$$\Omega_k{}^3 \prod_i \left(1 - \frac{x_{ik}}{\Omega_k}\right) > \Omega_l{}^3 \prod_i \left(1 - \frac{x_{jl}}{\Omega_l}\right) \tag{B2}$$

It is clear from this equation that for similar values of Ω, the largest is the one with all x's equal to zero, which is the XOR-like case. Thus, if there is an XOR-like dependence present in the data it will be at the top of the symmetric delta scoring list.

8.3. *Appendix C. A matrix formulation of the three-way dependence relations*

The set of equations in 9a can be expressed simply in matrix form by defining the vectors:

$$\vec{\Omega} \equiv \begin{bmatrix} \Omega \\ \Omega \\ \Omega \end{bmatrix}, \vec{\Delta} \equiv \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \end{bmatrix}, \vec{\mu} \equiv \begin{bmatrix} \mathrm{I}_{12} \\ \mathrm{I}_{13} \\ \mathrm{I}_{23} \end{bmatrix} \tag{C1}$$

All of the components of these vectors are non-negative, and for the moment we are not assuming that $I_{12} = 0$, so this holds for the general case. Then we have a matrix relation,

$$\vec{\Omega} = \vec{\Delta} + \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \vec{\mu} \tag{C2}$$

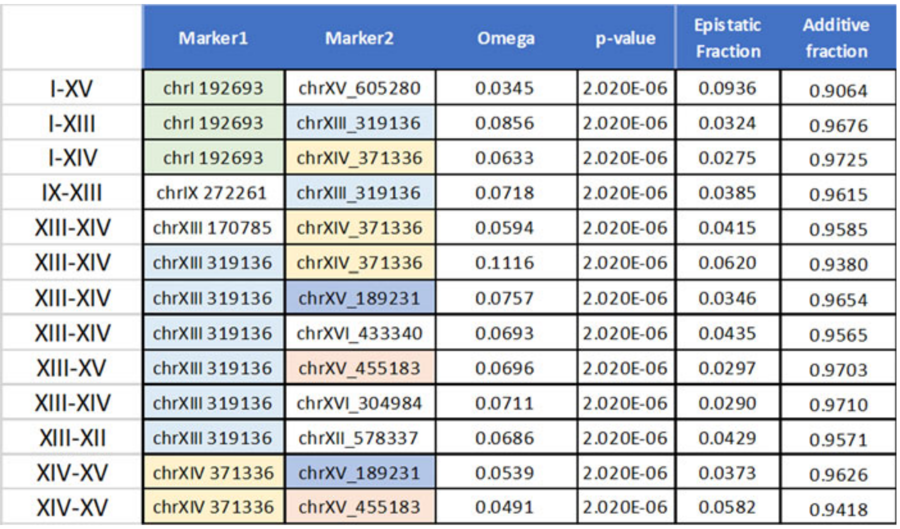
The functions are thus confined to the landscape in only the non-negative sector (all coordinates are ≥ 0). The matrix $\overline{\overline{M}} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, has the inverse, $\overline{\overline{M}}^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{bmatrix}$, and writing the expressions for the other two vectors gives these simple equations

$$\vec{\mu} = \overline{\overline{M}}^{-1} \left(\vec{\Omega} - \vec{\Delta} \right) \tag{C3}$$

Thus, the three-variable relations reduce to a simple equation in matrix form.

8.4. *Appendix D. Interactions in the yeast data (Bloom et al., 2015)*

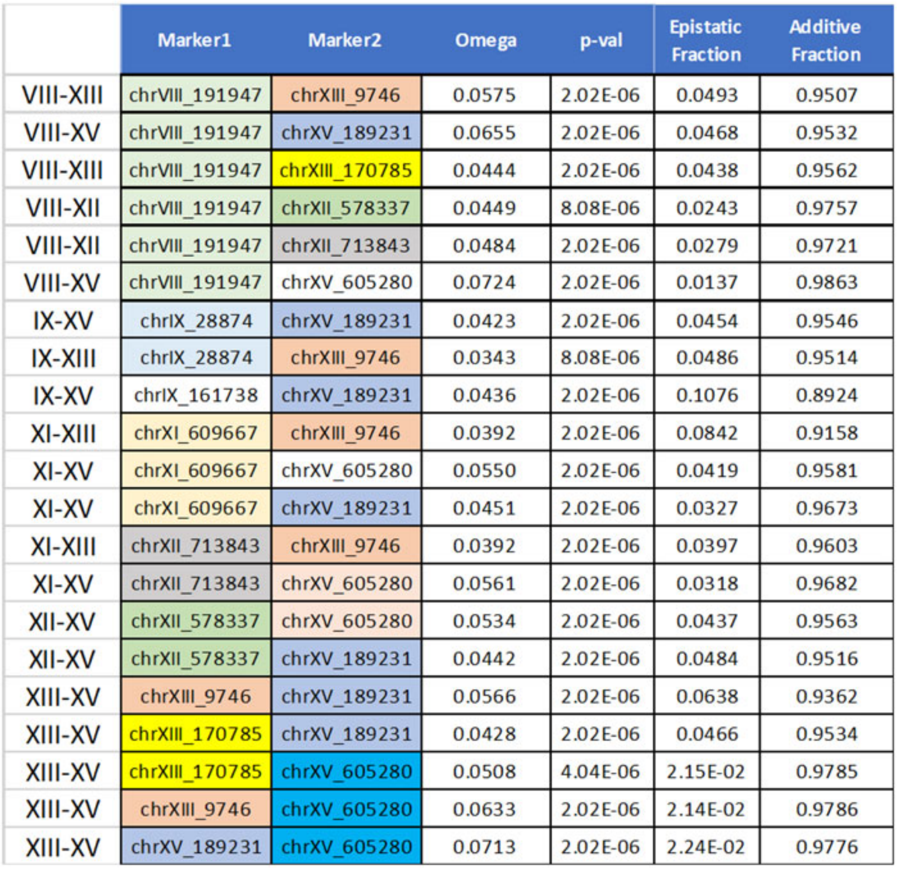
Appendix Table D1. The Interchromosomal Interactions for the Neomycin Phenotype, Exhibiting the Markers (Identical Markers Are Color-Coded for Better Visualization)



	Marker1	Marker2	Omega	p-value	Epistatic Fraction	Additive fraction
I-XV	chrI 192693	chrXV_605280	0.0345	2.020E-06	0.0936	0.9064
I-XIII	chrI 192693	chrXIII_319136	0.0856	2.020E-06	0.0324	0.9676
I-XIV	chrI 192693	chrXIV_371336	0.0633	2.020E-06	0.0275	0.9725
IX-XIII	chrIX 272261	chrXIII_319136	0.0718	2.020E-06	0.0385	0.9615
XIII-XIV	chrXIII 170785	chrXIV_371336	0.0594	2.020E-06	0.0415	0.9585
XIII-XIV	chrXIII 319136	chrXIV_371336	0.1116	2.020E-06	0.0620	0.9380
XIII-XIV	chrXIII 319136	chrXV_189231	0.0757	2.020E-06	0.0346	0.9654
XIII-XIV	chrXIII 319136	chrXVI_433340	0.0693	2.020E-06	0.0435	0.9565
XIII-XV	chrXIII 319136	chrXV_455183	0.0696	2.020E-06	0.0297	0.9703
XIII-XIV	chrXIII 319136	chrXVI_304984	0.0711	2.020E-06	0.0290	0.9710
XIII-XII	chrXIII 319136	chrXII_578337	0.0686	2.020E-06	0.0429	0.9571
XIV-XV	chrXIV 371336	chrXV_189231	0.0539	2.020E-06	0.0373	0.9626
XIV-XV	chrXIV 371336	chrXV_455183	0.0491	2.020E-06	0.0582	0.9418

The *p*-values, calculated from the permutation of the data, are approximately the same.

Appendix Table D2. The Interchromosomal Interactions for the Copper Sulfate Phenotype, Exhibiting the Markers (Identical Markers Are Color-Coded for Better Visualization)



	Marker1	Marker2	Omega	p-val	Epistatic Fraction	Additive Fraction
VIII-XIII	chrVIII_191947	chrXIII_9746	0.0575	2.02E-06	0.0493	0.9507
VIII-XV	chrVIII_191947	chrXV_189231	0.0655	2.02E-06	0.0468	0.9532
VIII-XIII	chrVIII_191947	chrXIII_170785	0.0444	2.02E-06	0.0438	0.9562
VIII-XII	chrVIII_191947	chrXII_578337	0.0449	8.08E-06	0.0243	0.9757
VIII-XII	chrVIII_191947	chrXII_713843	0.0484	2.02E-06	0.0279	0.9721
VIII-XV	chrVIII_191947	chrXV_605280	0.0724	2.02E-06	0.0137	0.9863
IX-XV	chrIX_28874	chrXV_189231	0.0423	2.02E-06	0.0454	0.9546
IX-XIII	chrIX_28874	chrXIII_9746	0.0343	8.08E-06	0.0486	0.9514
IX-XV	chrIX_161738	chrXV_189231	0.0436	2.02E-06	0.1076	0.8924
XI-XIII	chrXI_609667	chrXIII_9746	0.0392	2.02E-06	0.0842	0.9158
XI-XV	chrXI_609667	chrXV_605280	0.0550	2.02E-06	0.0419	0.9581
XI-XV	chrXI_609667	chrXV_189231	0.0451	2.02E-06	0.0327	0.9673
XI-XIII	chrXII_713843	chrXIII_9746	0.0392	2.02E-06	0.0397	0.9603
XI-XV	chrXII_713843	chrXV_605280	0.0561	2.02E-06	0.0318	0.9682
XII-XV	chrXII_578337	chrXV_605280	0.0534	2.02E-06	0.0437	0.9563
XII-XV	chrXII_578337	chrXV_189231	0.0442	2.02E-06	0.0484	0.9516
XIII-XV	chrXIII_9746	chrXV_189231	0.0566	2.02E-06	0.0638	0.9362
XIII-XV	chrXIII_170785	chrXV_189231	0.0428	2.02E-06	0.0466	0.9534
XIII-XV	chrXIII_170785	chrXV_605280	0.0508	4.04E-06	2.15E-02	0.9785
XIII-XV	chrXIII_9746	chrXV_605280	0.0633	2.02E-06	2.14E-02	0.9786
XIII-XV	chrXV_189231	chrXV_605280	0.0713	2.02E-06	2.24E-02	0.9776

The *p*-values, calculated from the permutation of the data, are approximately the same.

8.5. Appendix E. Estimates of entropies from data

There is a large literature on the estimation of entropies and mutual information that dates from shortly after Shannon's first articles on information theory. This issue needs to be addressed for any application of information theory, including this one, since the data we deal with are used to calculate measures entirely dependent on estimated entropies. It is easy to see intuitively how small-sample numbers can bias entropy estimations. Consider that we have discrete categories or alphabets for a variable. If there are only a few samples, the principal errors will likely be in underestimating the number of occurrences of categories. This, of course, will cause an underestimate of the entropy of the variable. As the sample numbers grow, this bias is reduced and the studies have addressed the questions: by how much, possible corrections, the estimation of error distributions, and so on (Nemenman et al., 2002). Good summaries of the literature and main issues can be found in Paninski (2003), Han et al. (2015), and Verdu (2019).

A good feel for the quantitative nature of the estimation problem can be obtained by considering an early formulation of a correction for the estimate of entropy from discrete data. The Miller-Madow bias correction for the maximum likelihood estimator of the entropy of a variable, well discussed in Paninski (2003), is given by the formula

$$H(pN) - MLE(pN) \approx \frac{m-1}{N} \tag{E1}$$

where $H(pN)$ is the true entropy, $MLE(pN)$ is the maximum likelihood estimate, m the number of letters in the alphabet with nonzero occurrence in the data, and N is the total number of samples for the variable. While this correction is not exact, and there is much more to the general problem, it points to the central issue, which is the degree to which the data properly represent the distribution of the variable. In our case where the alphabet size is small (for a diploid genotype genetic variable $m = 3$) and N is orders of magnitude greater, the Miller-Madow correction is therefore very small. This means, in general, that the entropy estimates are a good reflection of the true entropies using, as we do here, a simple, so-called naive estimate of the probabilities to calculate the entropies. Nonetheless, if the sample numbers decrease or the phenotype resolution increases in some cases, we need to be very aware of the errors in the estimation of the entropies as they affect all information measures.

AUTHOR CONTRIBUTIONS

D.J.G. conceived of the project, D.J.G., N.A.S., and J.K.-G. provided and developed key ideas for specific parts of the formalism and article sections. L.U. and J.K.-G. carried out calculations on the yeast data. D.J.G. wrote the first draft of the article and D.J.G., N.A.S., and J.K.-G. further edited and revised the article.

ACKNOWLEDGMENTS

We thank Joe Nadeau and Aimee Dudley for stimulating discussions, Greg Carter and Philip Welkoff for useful comments on an early version of the article, and Elaine Skeffington for expert copy and format editing of the article. None of the views expressed here should be attributed to the funding organizations. Neither did they play any role in the decision to publish or in the process of publication.

AUTHOR DISCLOSURE STATEMENT

The authors declare they have no competing financial interests.

FUNDING INFORMATION

This work was made possible by partial support from the, NSF, National Science Foundation, EAGER grant (IIS-1340619), by the Pacific Northwest Research Institute, and the Bill and Melinda Gates Foundation. In the early stages of this work, the Luxembourg Centre for Systems Biomedicine (LCSB) of the University of Luxembourg also provided significant partial support.

REFERENCES

Arboleda-Velasquez, J.F., Lopera, F., O'Hare, M., et al. 2019. Resistance to autosomal dominant Alzheimer's disease in *APO3* Christchurch homozygote: A case report. *Nat. Med.* 25, 1680–1683.

Bell, A.J. 2003. The co-information lattice. In ICA 2003, Nara, Japan, 921–926.

Bertschinger, N., Rauh, J., Olbrich, E., et al. 2012. Shared Information: New insights and problems in decomposing information in complex systems, 251–269. Proceedings of the ECCS. Springer, Cham.

Bloom, J.S., Kotenko, I., Sadhu, M.J., et al. 2015. Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. *Nat. Commun.* 6, 8712.

Churchill, G.A., and Doerge, R.W. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971.

Coutinho, A.M., Sousa, I., Martins, M., et al. 2007. Evidence for epistasis between SLC6A4 and ITGB3 in autism etiology and in determination of platelet serotonin levels. *Hum. Genet.* 121, 243–256.

Crow, J.F. 2001. Shannon's brief foray into genetics. *Genetics* 159, 915–917.

Crow, J.F. 2010. On epistasis: Why it is unimportant in polygenic directional selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365:1241–1244.

Eichler, E.E., Flint, J., Gibson, et al. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* 11, 446–450.

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52, 399–433.

Frieden, B.R., Plastino, A., and Soffer, B.H. 2001. Population genetics from an information perspective. *J. Theor. Biol.* 208, 49–64.

Galas, D.J., Dewey, G., Kunert-Graf, J., et al. 2017. Expansion of the Kullback-Leibler divergence, and a new class of information metrics. *Axioms* 6, 8.

Galas, D.J., Nykter, M., Carter, G.W., et al. 2010. Biological information as set based complexity. *IEEE Trans. Inf. Theor.* 56, 667–677.

Galas, D.J., Sakhanenko, N.A., Skupin, A., et al. 2014. Describing the complexity of systems: Multi-variable ''set complexity'' and the information basis of systems biology. *J. Comput. Biol.* 21, 118–140.

Gilbert-Diamond, D., and Moore, J.H. 2011. Analysis of gene-gene interactions. *Curr. Protoc. Hum. Genet.* 70, 1.14.1 – 1.14.12.

Gregersen, J.W., Kranc, K.R., Ke, X., et al. 2006. Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature* 443, 574–577.

Han, T.S. 1980. Multiple Mutual Information and multiple interactions in frequency data. *Inf. Control* 46, 26–45.

Han, Y., Jiao, J., and Weissman, T. 2015. Adaptive estimation of Shannon entropy, 1372–1376. Proceedings of IEEE International Symposium on Information Theory (ISIT), Hong kong.

Hill, W.G., Goddard, M.E., and Visscher, P.M. 2014. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* 4, e1000008.

Huang, W., and Mackay, T.F. 2016. The genetic architecture of quantitative traits cannot be inferred from variance component analysis. *PLoS Genet.* 12, e1006421.

Jaynes, E.T. 1957. Information theory and statistical mechanics. *Phys. Rev.* 106, 620–630.

Jaynes, E.T. 2005. On the rationale of maximum-entropy methods. *Proc. IEEE* 70, 939–952.

Lstiburek, M., Bittner, V., Hodge, G.R., et al. 2018. Estimating realized heritability in panmictic populations. *Genetics* 208, 89–95.

Mackay, T.F. 2014. Epistasis and quantitative traits: Using model organisms to study gene-gene interactions. *Nat. Rev. Genet.* 15, 22–33.

McGill, W.J. 1954. Multivariate information transmission. *Psychometrika* 19, 97–116.

Moran, P.A.P. 1961. Entropy, Markov processes and Boltzmann's H-theorem. *Math Proc. Cambridge Philos. Soc.* 57, 833–842.

Nelson, R.M., Petterson, M.E., and Carlborg, O. 2013. A century after Fisher: Time for a new paradigm in quantitative genetics. *Trends Genet.* 29, 669–676.

Nemenman, I., Shafee, F., and Bialek,W. 2002. Entropy and inference, revisited. *In* Dietterich, T.G., Becker, S., and Ghahramani, Z. eds. *Advances in Neural Information Processing Systems 14*. MIT Press, Cambridge, MA.

Paninski, L. 2003. Estimation of entropy and mutual information. *Neural Comput.* 15, 1191–1253.

Phillips, P.C., and Johnson, N.A. 1998. The population genetics of synthetic lethals. *Genetics* 150, 449–458.

Phillips, P.C. 2008. Epistasis—The essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9, 855–867.

Ridge, P.G., Karch, C.M., Hus, S., et al. 2017. Linkage, whole genome sequence, and biological data implicate variants in *RAB10* in Alzheimer's disease resilience. *Genome Med.* 9, 100.

Sakhanenko, N.A., and Galas, D.J. 2015. Biological data analysis as an information theory problem: Multivariable dependence measures and the Shadows algorithm. *J. Comput. Biol.* 22, 1005–1024.

Sakhanenko, N.A., and Galas, D.J. 2019. Symmetries among multivariate information measures explored using Möbius operators. *Entropy* 21, 88.

Sakhanenko, N.A., Kunert-Graf, J., and Galas, D.J. 2017. The information content of discrete functions and their application to genetic data analysis. *J. Comput. Biol.* 24, 1153–1178.

Shannon, C.E. 1940. An algebra for theoretical genetics [Ph.D. dissertation]. MIT (Department of Mathematics), Cambridge, MA.

Shannon, C.E. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.

Ting, H.K. 1962. On the amount of information. *Theor. Probab. Appl.* 7, 439–444.

Tyler, A.L., Donahue, L.R., Churchill, G.A., et al. 2016. Weak epistasis generally stabilizes phenotypes in a mouse intercross. *PLoS Genet.* 12, e1005805.

Verdu, S. 2019. Empirical estimation of information measures: A literature guide. *Entropy* 21, 720.

Watanabe, S. 1960. Information theoretic analysis of multivariate correlation. *IBM J. Res. Dev.* 4, 66–82.

Watterson, G.A. 1962. Some theoretical aspects of diffusion theory in population genetics. *Ann. Math. Stat.* 33, 939–957.

Wiltshire, S., Bell, J.T., Groves, C.J., et al. 2006. Epistasis between type 2 diabetes susceptibility loci on chromosomes 2q21–25 and 10q23–26 in Northern Europeans. *Ann. Hum. Genet.* 70, 726–737.

Wright, S.G. 1926. A frequency curve adapted to variation in percentage occurrence. *J. Am. Stat. Assoc.* 21, 162–178.

Address correspondence to:
Dr. David J. Galas
Pacific Northwest Research Institute
720 Broadway
Seattle, WA 98122
USA

E-mail: dgalas@pnri.org