



Annual Review of Genomics and Human Genetics
The Causes and Consequences
of Genetic Interactions
(Epistasis)

Júlia Domingo,^{1,*} Pablo Baeza-Centurion,^{1,*}
and Ben Lehner^{1,2,3}

¹Systems Biology Program, Centre for Genomic Regulation, Barcelona Institute of Science and Technology, 08003 Barcelona, Spain; email: julia.domingo@crg.eu, pablo.baeza@crg.eu, ben.lehner@crg.eu

²Universitat Pompeu Fabra, 08003 Barcelona, Spain

³Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

Annu. Rev. Genom. Hum. Genet. 2019.
20:17.1–17.28

The *Annual Review of Genomics and Human Genetics*
is online at genom.annualreviews.org

<https://doi.org/10.1146/annurev-genom-083118-014857>

Copyright © 2019 by Annual Reviews.
All rights reserved

*These authors contributed equally to this article

Keywords

epistasis, genetic interactions

Abstract

The same mutation can have different effects in different individuals. One important reason for this is that the outcome of a mutation can depend on the genetic context in which it occurs. This dependency is known as epistasis. In recent years, there has been a concerted effort to quantify the extent of pairwise and higher-order genetic interactions between mutations through deep mutagenesis of proteins and RNAs. This research has revealed two major components of epistasis: nonspecific genetic interactions caused by nonlinearities in genotype-to-phenotype maps, and specific interactions between particular mutations. Here, we provide an overview of our current understanding of the mechanisms causing epistasis at the molecular level, the consequences of genetic interactions for evolution and genetic prediction, and the applications of epistasis for understanding biology and determining macromolecular structures.



Intergenic epistasis: interaction between mutations in different genes (as opposed to intragenic epistasis, which occurs between mutations within the same gene)

INTRODUCTION

A mutation will not always have the same consequences in different individuals. For instance, a mutation that causes a disease in one individual may have no effect in another (22, 23, 65). There are several reasons for this. One is the environment: Diet, pathogens, temperature, and other risk factors vary among individuals and can change the effects of mutations. Another cause is stochastic processes, such as fluctuations and errors in gene expression (21, 24, 37, 122). However, the effect of a mutation may also vary across individuals because of additional genetic variation. Such dependencies on genetic background or context can cause phenotypic diversity within individuals and, on evolutionary timescales, incompatibilities between different species (26, 72, 120). For instance, some mutations that cause diseases in humans are fixed in the genomes of other species without any apparent deleterious consequences (19, 42, 62).

This dependency of the effects of mutations on the genetic background in which they occur is known as epistasis or genetic interaction. The concept of epistasis was introduced by Bateson (14) in the early twentieth century to describe the effect of a genetic variant that masked the effect of another. Fisher (39) defined epistasis as the statistical deviation from the additive combination of two loci in their effects on a phenotype. Since then, the term epistasis has been used with many related meanings (89, 114), although it is most commonly defined as the deviation from the expected outcome when combining mutations. For fitness, the expected outcome is usually taken as the multiplicative or log-additive combination of the individual mutation effects (89, 114). Epistatic interactions can be divided into different classes (115, 156) depending on whether the outcome is better (positive epistasis) or worse (negative epistasis) than expected, whether only the strength of a mutation's effect changes (magnitude epistasis) or also its direction of effect (sign epistasis), whether the interaction involves two (pairwise epistasis) or more (higher-order epistasis) mutations, or whether the interaction is particular to a mutation combination (specific epistasis) or completely predictable from knowledge of the phenotypic effects of the mutations without knowledge of their identity (nonspecific or general epistasis) (**Figure 1**).

Although epistasis has been studied for decades using experimental (118, 138, 154) and computational (62, 119) approaches, in recent years there has been a concerted effort to experimentally quantify genetic interactions in a high-throughput manner (5). This has allowed us to evaluate the abundance of epistasis within and between genes and better understand the underlying molecular mechanisms. The aim of this review is to offer an overview of how new experimental approaches to comprehensively quantify the effects of mutations have improved our understanding of the prevalence, causes, implications, and applications of epistasis.

THE PREVALENCE OF EPISTASIS WITHIN AND BETWEEN GENES

Epistasis between genes has historically been exploited by geneticists and developmental biologists to disentangle the order in which genes act in linear pathways (45, 145, 148). An example experiment involved combining a null allele of a gene essential for the activity of a pathway (so the null allele blocked the pathway) with a null allele of a gene that negatively affected the activity of the pathway (so the null allele enhanced the activity of the pathway), with the reasoning that if the essential gene was upstream of the other gene, the pathway would remain active. Genome screening projects with libraries of single and double gene deletions, inhibitions, or hypomorphic alleles (**Figure 2a**) have revealed that intergenic epistasis is abundant in different organisms, including yeast, worms, and humans (28, 38, 56, 77, 78, 81, 125). For example, the generation of 23 million double-knockout gene combinations encompassing 5,416 different genes of budding yeast identified nearly a million interactions (~4% of tested pairs) when screening for growth (colony size) (28). These results not only demonstrate the prevalence of pairwise genetic

17.2 Domingo • Baeza-Centurion • Lehner

Review in Advance first posted on
May 13, 2019. (Changes may still
occur before final publication.)



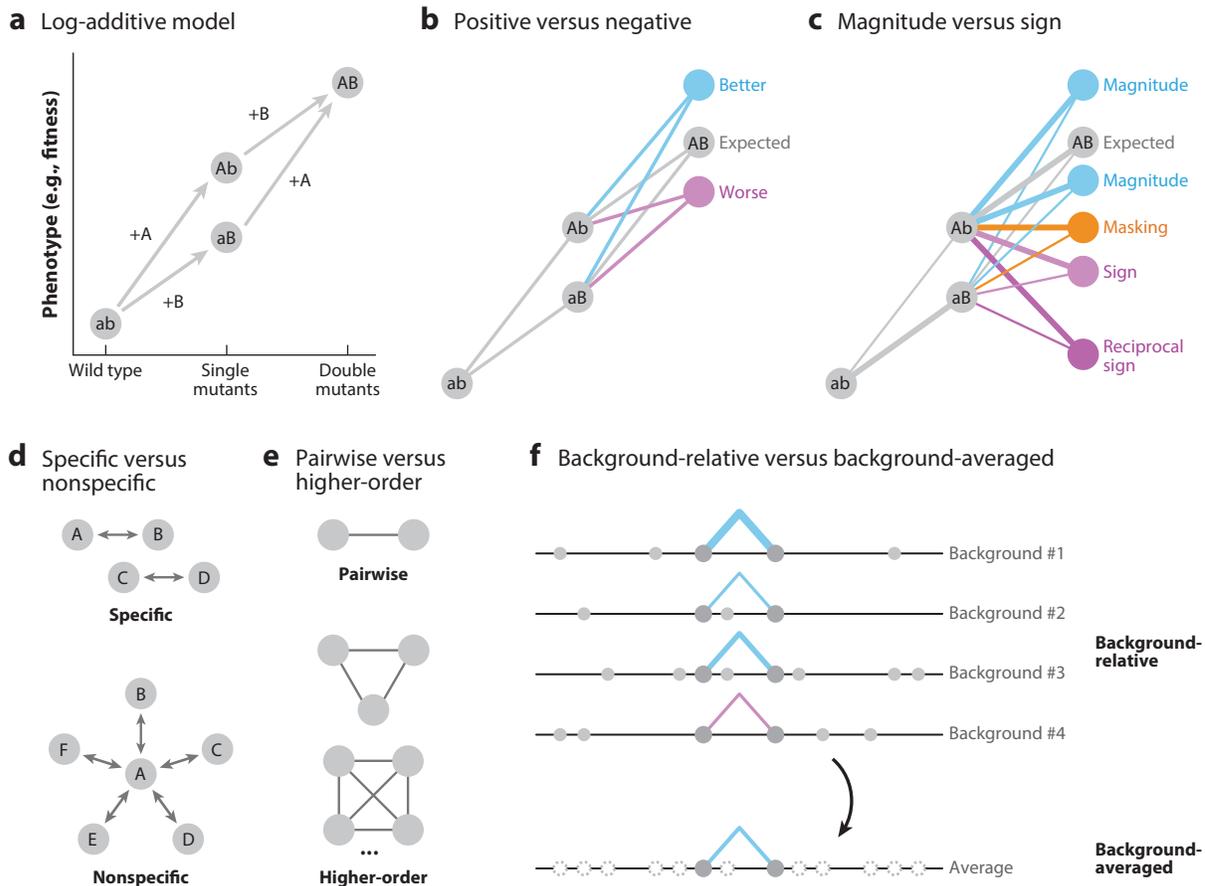


Figure 1

Classes of epistasis. (a) The log-additive model for how mutations combine. (b) Positive versus negative epistasis. Positive epistasis occurs when the effect of a mutation in a given genetic background is better (or fitter) than expected, and negative epistasis occurs when the effect of a mutation is worse (or less fit) than expected. The specific case of a beneficial mutation that becomes less beneficial is referred to as diminishing-returns epistasis. (c) Magnitude versus sign epistasis. Magnitude epistasis occurs when the magnitude but not the direction of the effect of a mutation changes in certain genetic backgrounds. If the effect of a mutation disappears in a given genetic background, that mutation is said to undergo masking epistasis, and if the direction of the effect changes, the mutation undergoes sign epistasis. A special case of sign epistasis, reciprocal sign epistasis, occurs when the direction of effect of both mutations changes. (d) Specific versus nonspecific epistasis. If the interaction between two mutations depends on the identity of the mutations involved, it is a specific interaction. If the interaction depends only on the magnitude of the effect of both mutations, irrespective of the exact identity of the mutations involved, it is a nonspecific interaction. In the case of a nonspecific interaction, the same mutation can interact with many other mutations. (e) Pairwise versus higher-order epistasis. Epistatic interactions can involve two (pairwise epistasis) or more (higher-order epistasis) mutations. (f) Background-relative versus background-averaged epistasis. An interaction between two mutations may differ depending on the genetic background. The interaction quantified in each background is referred to as background-relative epistasis. Background-averaged epistasis quantifies the average interaction between two mutations across a set of genetic backgrounds.

interactions but also provide a resource for understanding the organization of the cell. Genes with closely related functions tend to have similar genetic interaction profiles, allowing the identification of genes belonging to related pathways, processes, and compartments (28).

Gene deletions are, however, rare (140); variation between individuals typically consists of point mutations, single-nucleotide insertions or deletions, gene duplications, and genomic

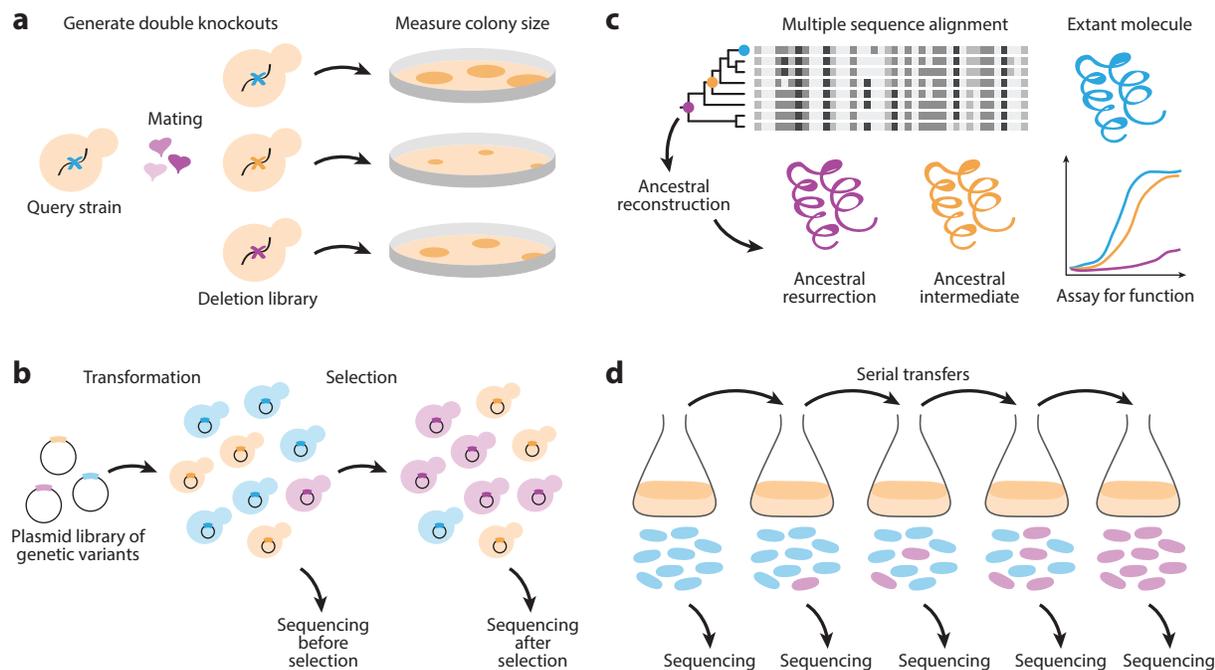


Figure 2

Experimental approaches to studying epistasis. (a) Synthetic lethal analysis. A haploid yeast strain with a deleted gene is mated with other haploid strains with other genes knocked out to build a double-knockout diploid strain. The colony size of the double knockout is a measure of the fitness of the yeast when both genes are knocked out. In the case of essential genes, hypomorphic alleles that cause a partial loss of gene function are used. (b) Deep mutational scan. A library of mutant variants of a gene necessary for cell proliferation is transformed into cells (or, alternatively, the variants may be integrated into the genome). The cells are then subjected to a set of conditions (selection conditions) under which they are expected to grow. Sequencing the plasmids from a sample of the cell population before selection and from a sample after selection and comparing the relative frequencies of each mutant variant in both samples provides an estimate of the effect of each variant on fitness. Information about epistasis can be obtained by measuring the fitness of every variant alone and in combination with other variants. (c) Ancestral reconstruction and resurrection. A multiple sequence alignment with the sequences of extant species allows the reconstruction of the sequence of the ancestral state, as well as that of any intermediate states between the ancestor and the present-day sequence. These reconstructed sequences are then used to synthesize (resurrect) the ancestral molecules, which all differ in the specific mutations harbored and which are then assayed for function and compared with each other. (d) Experimental evolution. A population is allowed to grow in a container with a specific set of selection conditions until it reaches a given size. A sample of this population is then taken from the first container and transferred into another one. This process is repeated many times at regular intervals. Over time, mutations are acquired that allow the members of the population to thrive under the experimental conditions used. The population can be sequenced at each time point or at the end of the experiment. Information about epistasis can be obtained by measuring the fitness of individual mutations as well as their combinations.

rearrangements, which have diverse effects on gene activity that are difficult to predict (140). Emergent techniques such as deep mutational scans (**Figure 2b**), which measure the effects of thousands of mutations on gene function with one experiment, have allowed researchers to test for the abundance of intragenic epistasis. These experiments have quantified interactions by mutating proteins (1, 6, 11, 32, 41, 49, 93, 97, 105, 111, 136, 144, 157) and RNAs (34, 46, 84, 121). Systematic mutagenesis has also been applied to regions of genes regulating transcription (113, 127) or alternative splicing (9, 18, 63, 66). For example, a study looking at the effects of mutations in the RNA recognition motif (RRM) domain of the yeast poly(A)-binding protein (PAB1) found that almost 20% of double mutants in this domain displayed some form of epistasis (97). The proportion of interacting mutation combinations varies extensively across studies, with

differences in analysis methods and experimental design making direct comparisons between studies difficult. Quantitative comparison of the extent of epistasis across molecules and systems remains an important challenge for the future.

MOLECULAR CAUSES OF EPISTASIS

Understanding what causes epistasis is important in order to predict phenotypic variation from changes in genotype. Despite the progress made in quantifying genetic interactions, our knowledge of their underlying molecular mechanisms remains limited (79). Large-scale data sets generated during the last few years have provided an opportunity to better understand the molecular causes of epistasis. Here, we distinguish two general classes of interaction: nonspecific epistasis, which arises from the nonlinearities in genotype–phenotype maps and applies to all mutation combinations, and specific epistasis, which occurs only between particular sets of mutations (Figure 1*d*).

MECHANISMS OF NONSPECIFIC EPISTASIS

If the relationship between the biophysical effect of a mutation and the measured phenotype is linear, then two mutations behaving additively at the level of their biophysical effects will also behave additively at the phenotypic level (Figure 3*a*). However, they will appear to behave non-additively if the relationship between the underlying additive trait and the measured phenotype is not linear (Figure 3*b*). The unexpected outcomes of mutations driven by a nonlinearity in the genotype-to-phenotype map are known as nonspecific interactions. They have also been referred to as global epistasis (108), unidimensional epistasis, and the fitness potential (73).

Predicting Nonspecific Interactions Without Understanding Causal Mechanisms

Nonspecific genetic interactions can be predicted from the effects of the individual mutations without knowing the identity of the mutations involved. Different methods have been used to deduce the shapes of these general nonlinearities even when the underlying causal mechanisms are unknown. A simple approach to infer the shape of this nonlinearity is to plot the observed effect of double mutations versus the effects of each of the two single mutations, with a running median surface or loess regression describing the overall relationship between the single- and double-mutation effects, and specific interactions quantified as the residuals from these best-fit surfaces (137) (Figure 4*a*, left). Another method involves plotting the observed versus the expected double-mutant effects by assuming mutations behave additively and fitting a power function (132) (Figure 4*a*, right). A maximum likelihood approach has also been described that extracts the underlying, unobserved, additive traits and fits an I-spline to the relationship between the phenotype and this unobserved trait (108) (Figure 4*a*, center). Finally, a study of the genotype-to-phenotype landscape of GFP (136) modeled the relationship between the underlying biophysical trait and the measured phenotype (fluorescence) as a sigmoid function that was later refined using a neural network. Although these methods all provide insight into the general shapes of genotype-to-phenotype landscapes, they do not help researchers understand the origins of these nonlinearities. Recent studies have attempted to elucidate some of the causes of these general nonlinearities, which include the thermodynamics of protein folding and molecular interactions (32, 86, 105, 150, 158), cooperativity (17, 47, 86, 93, 100), mutually exclusive molecular competitions (9), enzyme kinetics (87, 146), feedback (68), and the balance between the costs and benefits of gene expression (25, 31).



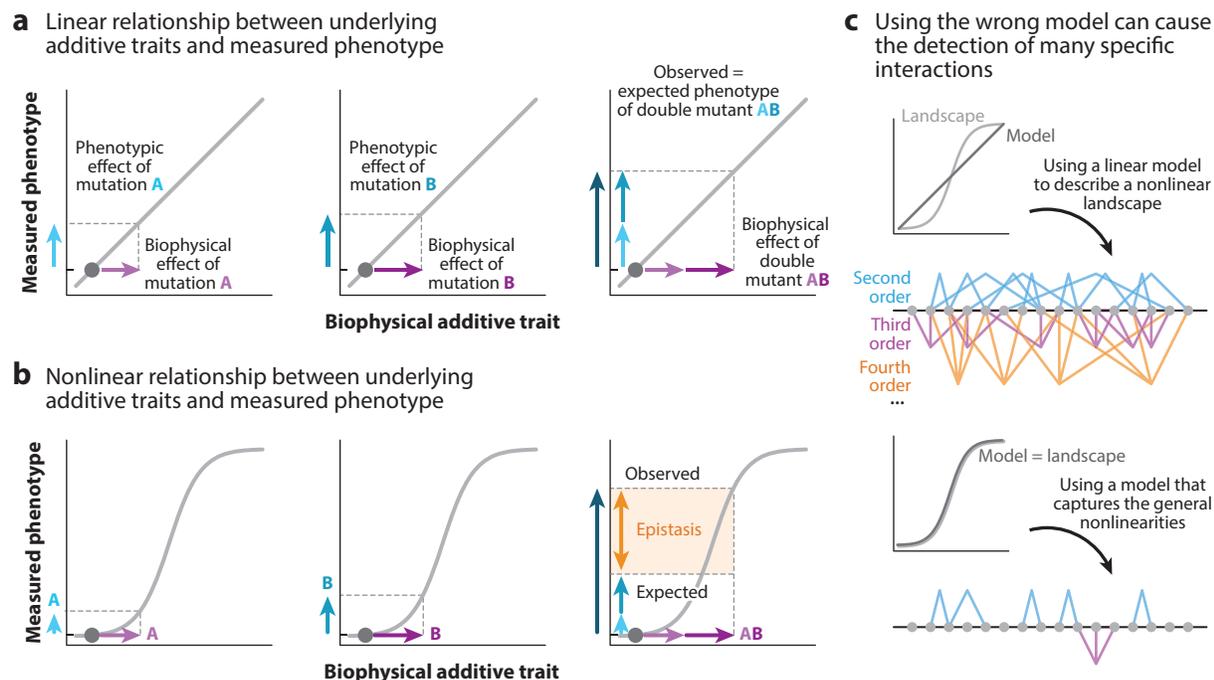


Figure 3

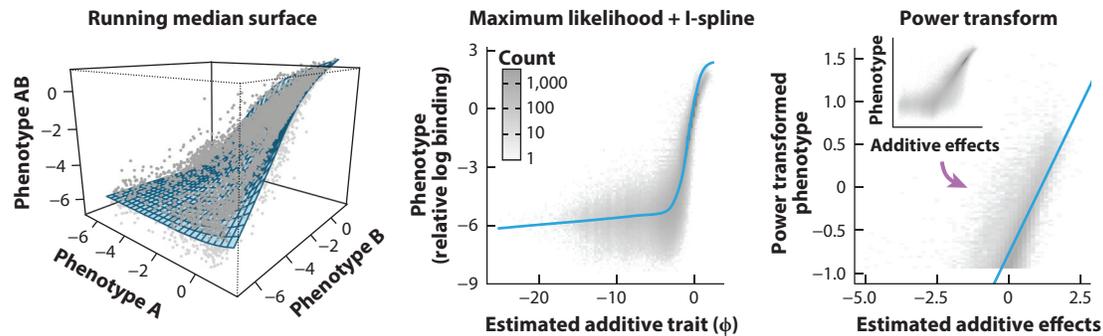
How a nonlinearity can lead to nonspecific epistasis. (a) Linear landscape. The effects of mutations A and B are additive in the underlying biophysical space and in the phenotypic space where their effects are measured. (b) Nonlinear landscape. The curvature in the function linking the underlying biophysical space (e.g., free energy of folding), where mutations have an additive effect, with the phenotype space, where mutation effects are measured means that mutations combine nonadditively. (c) General nonlinearities can cause the detection of spurious specific interactions. Using a linear model to describe a nonlinear genotype-to-phenotype landscape results in the detection of many specific interactions, including third-, fourth-, and higher-order terms, that attempt to describe this nonlinearity. If the null model for how mutations combine considers the general nonlinearity, fewer specific terms are detected.

Thermodynamics of Protein Folding and Molecular Interactions

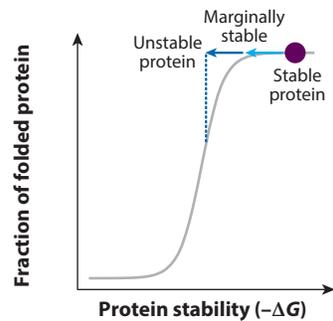
The relationship between the stability—the free energy of folding, ΔG_F —of a protein and the fraction of the protein in its native folded state is sigmoidal (**Figure 4b**). When mutations affect ΔG_F but the assay used to estimate their effects is a higher-level property that depends on the amount of folded protein, then the effects of mutations will combine in a sigmoidal rather than a linear manner (150, 158). Wild-type proteins tend to be positioned on the upper plateau of this sigmoidal curve, meaning that mildly detrimental mutations can have little effect alone, an effect referred to as threshold robustness (150) (**Figure 4b**). However, adding a second mutation of similar small effect will have a larger impact on the fraction of folded protein because the protein's stability reservoir becomes exhausted and the protein no longer falls on the upper plateau of the sigmoid (105, 107) (**Figure 4b**). The thermodynamics of protein folding have been proposed to underlie much of the nonspecific epistasis detected in deep mutational scanning experiments of protein G domain B1 (GB1), where the effects of all single-amino-acid substitutions and their pairwise combinations were quantified (105, 107, 137).

A separate study used a protein fragment complementation assay in combination with deep sequencing to study the protein–protein interaction between the proteins FOS and JUN (32). Single-amino-acid substitutions were introduced into each of the proteins, alone or in

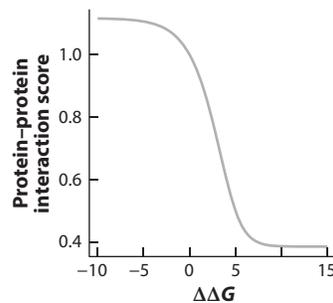
a Nonmechanistic methods of estimating global nonlinearities



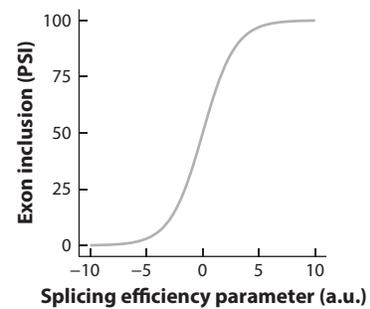
b Threshold robustness model



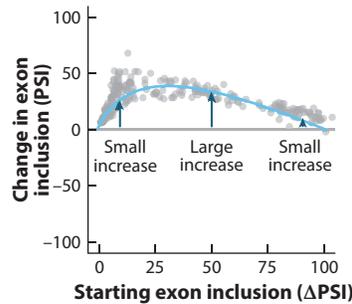
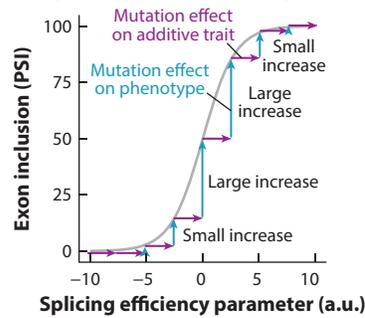
c Thermodynamics of a protein-protein interaction



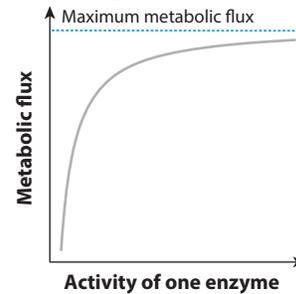
d Mutually exclusive splice site competition



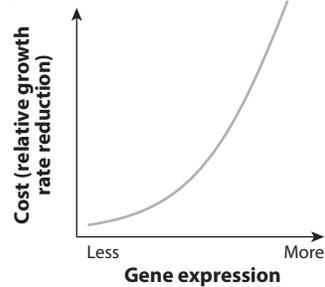
e Sigmoid generates scaling of mutation effects



f Saturating curve

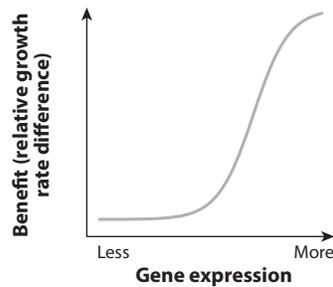


g Costs of gene expression

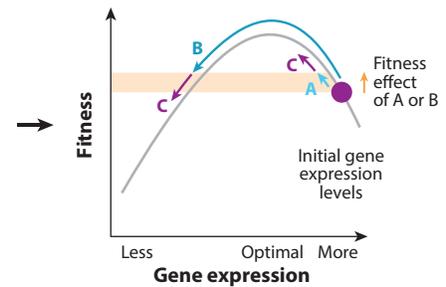


+

Benefits of gene expression



Peaked fitness landscape



(Caption appears on following page)

Figure 4 (Figure appears on preceding page)

Molecular mechanisms underlying nonspecific epistasis. (a) Different methods to estimate the global nonlinearity between the underlying additive trait and the observed phenotype. (Left) A running median surface describes the relationship between the phenotypes of the single mutants A and B and the double mutants AB. (Center) A maximum likelihood approach was developed to estimate the underlying additive traits of the genotype-to-phenotype landscape and an I-spline fit to describe the relationship between the observed phenotype and the underlying trait. (Right) A power transform can be used to linearize the relationship between the additive effects of the single mutants and the phenotype of the double mutants (the method is described in Reference 108). (b) Threshold robustness model of protein folding. The relationship between protein stability (the underlying biophysical space where mutations could have an effect) and the fraction of natively folded protein is sigmoidal (gray line). A stable protein lies at the top of this sigmoid, and introducing a destabilizing mutation (light blue arrow) would not significantly reduce the fraction of folded protein. However, after acquiring this mutation, the protein becomes only marginally stable, and the introduction of a second mutation (darker blue arrow) with a very similar effect on protein stability suddenly has a large effect on the fraction of folded protein (dotted blue line). (c) Nonlinearity introduced by the thermodynamics of a protein interaction. The relationship between the change in free energy introduced by a mutation (x axis) and the strength of a protein–protein interaction (y axis) is sigmoidal. (d) Nonlinearity introduced by mutually exclusive competition. The relationship between the efficiency of exon splicing (x axis) and the percentage of exon inclusion in the mature mRNA (y axis) is sigmoidal. (e) Scaling of mutation effects. The relationship between splicing efficiency (the underlying biophysical trait where mutations have an effect) and exon inclusion (the phenotypic space where the effects of mutations are measured) is sigmoidal, which means that mutations will have a maximum effect when introduced into an exon included at intermediate levels. The effects of mutations decrease when they are introduced into exons with very high or very low levels of inclusion. (f) Nonlinear relationship between metabolic flux and the activity of one enzyme in this pathway. The function saturates at maximum metabolic flux. (g) The costs and benefits of gene expression leading to a peaked fitness landscape. (Left) Expressing a gene has a cost due to, for example, energy consumption or competition for cellular resources. (Center) There are also benefits associated with the function carried out by the gene product in the cell. (Right) The combination of costs and benefits associated with gene expression leads to a peaked fitness landscape with an optimal gene expression level. The fitness of an organism with a gene expressed at levels above the optimum can be increased by introducing a mutation (A or B) that decreases the expression of that gene. Note that the same fitness can be obtained with two mutations causing very different reductions in gene expression levels. Introducing another mutation (C) that decreases gene expression will result in increased fitness if the protein has not yet reached the optimum with the first mutation (A), or in reduced fitness (sign epistasis) if the first mutation (B) had already passed the optimum. Abbreviation: PSI, percentage spliced in. Left and center subpanels of panel *a* adapted from References 137 and 132, respectively; panel *c* adapted from Reference 32; panels *d* and *e* adapted from Reference 9.

DeepPCA:

a high-throughput technique based on the protein fragment complementation assay that quantifies how mutations in two proteins combine to alter protein–protein interactions

Permissive mutation:

a mutation that allows the tolerance of another mutation that is otherwise deleterious

combination with a mutation in the other protein, and the effects on the strength of the interaction were estimated by measuring their effects on cell growth using a sequencing-based protein interaction assay called deepPCA (32). A multiplicative model for how mutations combine accounted for approximately 85% of the variance of the double-mutant effects, although systematic biases were present in the model predictions. For example, two mutations that only slightly destabilize the protein–protein interaction often lead to a weaker interaction than expected given the effect of the individual mutations. These biases imply the presence of nonspecific epistasis, which could be predicted using a three-parameter thermodynamic model based on the two-state equilibrium between the unfolded peptides and complex formation. This sigmoidal relationship could predict how pairs of mutations combine both between and within the individual proteins to alter complex formation (32) (Figure 4c).

Cooperativity

A second major mechanism behind nonspecific epistasis is the nonlinear relationship between protein concentration and bound ligand in cooperative binding. A typical example of cooperativity involves transcription factors binding to DNA (47, 99). Steroid receptors are a class of transcription factors that form homodimers and bind to sites known as steroid response elements or estrogen response elements. The evolution of an ancestral estrogen–response–element–binding receptor into a steroid–response–element–binding receptor involved three amino acid substitutions in the DNA recognition domain of the protein (93). However, these specificity-switching mutations were not tolerated except in the presence of an additional 11 permissive mutations that did not increase

steroid-response-element binding specificity but instead promoted the binding of both steroid response elements and estrogen response elements alike (93). Although these permissive substitutions were originally thought to increase the thermodynamic stability of the protein, as described in the previous section, reversible chemical denaturation experiments revealed that these mutations increased cooperativity between homodimer subunits instead of increasing protein stability. This experiment looked at cooperativity between homodimers, but synergistic epistasis arising due to the cooperative binding of different transcription factors has also been reported (17).

Multiple Nonlinearities Normally Connect Genotype to Phenotype

The examples of nonspecific epistasis described above focused on the analysis of molecular phenotypes one nonlinearity away from the layer where a mutation has its direct additive effect. This means that one nonlinear function is enough to describe the relationship between the measured and the underlying biophysical effect of a mutation. However, from transcription to RNA processing, translation, and protein folding and all the way up to protein activity and cellular fitness, there are many layers of biological organization where the effects of a mutation can be transformed. For many genes, nonspecific epistasis reflects multiple linked nonlinear functions. To address this concern and study how mutational effects propagate through different layers of biology, a deep mutational scan was performed on the DNA-binding domain of the phage lambda repressor (86), a transcription repressor. In this experiment, the target regulatory region was inserted upstream of a GFP reporter, and the decrease in GFP fluorescence was used to measure repressor activity. As described above, the concentration of natively folded protein is nonlinearly related to the folding energy of the protein (32, 150). However, also as described above, cooperativity in the recruitment of repressor dimers to their binding sites means that lambda repressor activity, dependent on the bound molecules, is itself nonlinearly related to the concentration of free lambda repressor (2). A model considering either one of these nonlinearities alone gives poor predictions of double-mutant phenotypes from the effects of the component single mutations. However, combining both nonlinearities gives good predictions (86). The combined model also accounts for why epistasis changes with different gene expression levels: Although changes in expression level normally have little effect on how mutations combine to alter the fraction of folded protein (the first nonlinearity), an increase in protein concentration alters how mutations combine to alter repression of the transcriptional target (the second nonlinearity).

Mutually Exclusive Competition

Mutually exclusive molecular competitions, in which two molecules compete to carry out the same function, are another potentially important contributor to nonspecific epistasis. This mechanism was recently described in a study of the effects of mutations on the inclusion of an alternative exon, *EAS* exon 6 (9). Exon inclusion was modeled as a competition between pairs of splice sites, resulting in a mutually exclusive outcome for each mRNA (the exon is either included or skipped). This model reveals a sigmoidal relationship between the efficiency of exon splicing (the underlying additive trait) and the final exon inclusion levels (the measured phenotypic trait) (Figure 4d). A consequence of this sigmoid is the nonmonotonic dependence of the effects of a mutation on the initial exon inclusion levels (Figure 4e). Naively, one might expect an inclusion-promoting mutation to have its largest effect in an exon with very low inclusion (which could in principle allow for large increases in inclusion) and its smallest effect in exons near 100% (inclusion cannot be above 100%). However, the largest mutation effects occurred in exons included at intermediate levels. This is because an exon with high or low levels of inclusion is

Synergistic epistasis: epistasis in which the effect on fitness of two mutations is more radical (more beneficial or more detrimental) than expected



near one of the horizontal asymptotes of the sigmoid, where the function is not steep and changes in splicing efficiency cause small changes in exon inclusion (**Figure 4e**). However, an exon with intermediate levels of inclusion is in the steep part of the sigmoid, where small changes in splicing efficiency cause large changes in exon inclusion. A similar nonmonotonic dependence of mutation outcome on the initial phenotype will occur for any system with a sigmoidal genotype–phenotype function. Although this molecular competition model was originally built to describe alternative splicing, it could be used to describe other processes involving a mutually exclusive molecular competition, such as signaling cascades involving alternative protein interaction partners (70). The prevalence of molecular competitions in biology suggests that mutually exclusive competition is an underappreciated contributor to nonspecific epistasis.

Nonlinear Expression–Fitness Functions

Another source of nonspecific epistasis arises in the nonlinear relationships between gene (protein) expression levels and fitness. Keren et al. (69) quantified the relationship between expression and growth rate for 81 genes in budding yeast and found that most genes have nonlinear expression–fitness functions, although the nature of this nonlinearity depends on the transcribed gene. For example, the expression of some genes is positively correlated with fitness, whereas for others, high levels of expression are detrimental. A third class of genes have peaked expression–fitness functions with an optimal expression level, above and below which fitness decreases. For most genes, the causes of these nonlinear functions are unknown, although they contribute nonspecific interactions that can be predicted even in the absence of mechanistic models (86).

One cause of nonlinear expression–fitness functions is metabolism. The flux of a metabolic pathway is not linearly related to the activity of an individual enzyme. Instead, it is linked by a concave function saturating at maximum metabolic flux (64) (**Figure 4f**). This is because enzymes do not act alone, but are kinetically linked to other enzymes through their substrates and products. Since there are usually many enzymes acting in the same metabolic pathway, the effect of changing the activity of any one of them often has a negligible effect on the overall flux. The consequences of this general nonlinearity were observed in an experiment carried out to study *Escherichia coli* isopropylmalate dehydrogenase (IMDH) (87). IMDH is involved in the biosynthesis of leucine and uses nicotinamide adenine dinucleotide as a coenzyme for catalysis. The authors created a genotype–phenotype landscape for coenzyme use by IMDH after introducing six substitutions in the coenzyme-binding pocket alone or in combination. When all genotypes were assayed for coenzyme usage, mutations behaved nonadditively at the level of bacterial fitness despite displaying an additive effect at the level of cofactor binding. Similarly, the effects of mutations reducing the affinity of β -lactamase for ampicillin have been proposed to be masked by the robustness conferred by the nonlinear relationship between fitness and enzyme activity (146).

Peaked Fitness Landscapes

An experiment examining the costs and benefits of Lac protein expression in *E. coli* (31) revealed that the cost of expressing the protein (the burden of transcribing and then translating the mRNA molecule) and the benefit (growth induced by the activity of the protein) combine to give a peaked expression–fitness function (**Figure 4g**). A peaked landscape was studied in an experiment where a plasmid containing genes essential for growth was transformed into bacteria (25). These genes had benefits (they were essential) and costs (the plasmid overexpressed them), and only a model that included the costs and benefits of protein expression could explain most (98%) of the variability in the data. Kemble et al. (68) recently built a data set of 1,369 genotypes to study the interactions

between mutations affecting the expression of two different genes, *araA* and *araB*, acting in the L-arabinose catabolism pathway in *E. coli*. Although this pathway can help *E. coli* grow in the presence of L-arabinose, the accumulation of its intermediate L-ribulose-5-phosphate is toxic to the cell. To understand how mutations combine, the authors built a mathematical model in which fitness depends on the expression of each of the two genes as well as the balance between the benefits of catabolic flux and the detrimental effects of intermediate toxicity. This is conceptually similar to the previously described model that balances the costs and benefits of gene expression, explaining why this model results in a peaked fitness landscape with optimal values for *araA* and *araB* gene expression.

Peaked landscapes provide a mechanism for sign epistasis (25, 68). Consider a gene expressed at levels above the optimum and two mutations, A and B, each increasing fitness (y axis in **Figure 4g**) by reducing gene expression (x axis). Mutation C, which also reduces gene expression, is beneficial in a background containing mutation A but deleterious in the presence of mutation B (**Figure 4g**). This is because mutation A reduces expression, bringing it closer to, but not reaching, the optimum. Thus, the addition of another expression-reducing mutation bringing the system even closer to the optimum is beneficial (**Figure 4g**). However, mutation B reduces gene expression past the optimum, and further reduction is deleterious. This is why the presence of a peaked fitness landscape means that, although two mutations may have similar effects on fitness, they can interact very differently with other mutations.

MOLECULAR MECHANISMS OF SPECIFIC EPISTASIS

In contrast to nonspecific epistasis, specific epistasis depends not only on the effect size of each mutation but also on the identity of the mutations involved. Specific epistasis, also known as structural epistasis (32), arises primarily through the distinctive effects of particular combinations of mutations—for instance, pairs of residues that contact each other in a folded protein or confer specificity for a particular ligand. The origins of specific epistasis are more diverse than nonspecific interactions and therefore are more difficult to predict. Specific and nonspecific epistasis are not mutually exclusive, but specific epistasis and the mechanisms that generate it become more apparent when the general nonlinear mapping between the measured phenotype and the underlying additive trait is taken into account (9, 32, 108, 132, 136). If not accounted for, nonspecific epistasis will result in the detection of many specific interactions—often including higher-order epistasis—that attempt to capture the nonlinear genotype–phenotype relationship (**Figure 3c**).

Specific Epistasis Due to the Three-Dimensional Structure of Molecules

An intuitive case of specific epistasis relates to the structure of molecules. Proteins and RNAs fold into defined structures due to physical interactions between their residues. Hydrophobic interactions, salt bridges, and hydrogen bonds occur between particular residues. When mutated, the disruption of these specific physical contacts can give rise to strong genetic interactions. An illustrative example is the strong positive epistasis generated by mutations at two positions that form a salt bridge in a protein (**Figure 5a**, top). Most individual substitutions will break the bridge and have a deleterious effect, especially if they introduce a repulsive electrostatic charge. However, a few specific mutations can restore the salt bridge by compensating for the newly introduced charge, keeping the structure intact and resulting in positive epistasis.

The idea that specific genetic interactions relate to energetic couplings and structural contacts in molecules is quite old, and biochemists have used double-mutant cycles to probe the folding and stability of proteins (57). This technique measures the change in free energy when two mutations



Figure 5 (Figure appears on preceding page)

Molecular mechanisms underlying specific epistasis. (a) Specific structural epistasis. (Top) A positive epistatic interaction between two residues that restore a salt in a protein structure. (Bottom) A structural interaction in the Fos–Jun complex (Protein Data Bank entry 1FOS). Dashed yellow lines represent polar interactions. (b) Partial correlations of epistasis values between pairs of positions in a protein to discriminate local from indirect contacts. (Left) The protein, showing three residues highlighted in different colors (two of them close in structure and one far away). (Right) Plotted epistasis scores between the three previously highlighted positions and the remaining positions in the protein. Positions that are close in three-dimensional space interact similarly with other positions. Partial correlations between interaction profiles are used to discriminate direct from indirect contacts. (c) Change in ligand specificity in the ancestral corticoid receptor driven by two interacting residues. (Left) Concentrations of the hormones aldosterone (A), deoxycorticosterone (D), and cortisol (C) required for half-maximal activation (EC₅₀) of the ancestral corticoid receptor (AncGR) in the presence or absence of the two mutations, S106P and L111Q. (Right) Structural rearrangement produced by the S106P and L111Q mutations in the ancestral receptor ligand-binding domain. The ancestral receptor is shown in green, and the ancestral receptor with the mutations is shown in yellow. (d) Epistasis generated by transcription factors that bind different DNA sequences, showing the three-dimensional binding-affinity landscapes of two different transcription factors for different DNA sequences. The *x* and *y* axes represent sequence space, and the *z* axis represents binding affinity. The DNA sequences with the highest binding affinity (optimal) are labeled on top of each peak. In the single-peaked landscape, mutations are additive at the level of free energy of binding, whereas in the two-peaked landscape, mutations display a strong epistatic behavior. (e) Mechanisms that generate epistasis in alternative splicing. (Top) Positive masking epistasis between two mutations in the exon, both of which create an overlapping and mutually exclusive RNA-binding motif, so the combination of the two mutations results in a smaller decrease in exon inclusion than that expected for the single-mutant combinations. (Bottom) Negative sign epistasis caused by one mutation breaking the binding site of a splicing repressor and the second one creating a new repressor binding site. Abbreviation: PSI, percentage spliced in. Bottom subpanel of panel *a* adapted from Reference 32; panel *e* adapted from Reference 9.

are introduced individually or in combination. Two residues are coupled when the change in free energy introduced by the double mutant differs from the sum of the changes in free energy of the single mutations alone. Although double-mutant cycles have proved useful to understanding how interactions between residues shape the structure and folding of proteins, only a small fraction of mutant combinations have been tested for any protein.

Recent deep mutational scans have shown that residues close in the primary sequence are enriched for epistasis. This reflects structural information about alpha helices or beta sheets of proteins (97, 105) or helical structures in RNAs (34). Similarly, residues close in three-dimensional space but not in the primary sequence are also enriched for epistasis. In RNAs, the most positive interactions occur when mutations restore base pairing (34, 46, 84, 121). In proteins, both positive epistasis and negative epistasis are enriched between positions close in three-dimensional space (97). For instance, a deep mutational scan of the RRM2 domain of the yeast PAB1 protein showed that approximately 17% of all significantly positive genetic interactions happened between residues less than 12 Å apart, and approximately 7% of all negatively interacting mutations were enriched for residues separated by 15 Å (97). Structural genetic interactions become more apparent when nonspecific epistasis is accounted for (32). In a systematic study of *cis* and *trans* genetic interactions between the two proto-oncogenes *FOS* and *JUN* (32), the epistasis that remained after nonspecific epistasis was accounted for was highly enriched among physically contacting and proximal residues (Figure 5a, bottom).

Nonetheless, many genetic interactions between and within molecules involve positions that are not in direct physical contact in the three-dimensional structure (long-range indirect interactions). This has been shown both by using multiple sequence alignments to analyze covariation between pairs of amino acids, reflecting their evolutionary dependence (48, 124), and by mutagenizing protein domains whose residues interact in cooperative networks to preserve protein function (94, 135). Amino acids group into protein sectors that include direct and allosterically transient interactions, so distinguishing direct from indirect structural interactions using genetic data alone is a challenging task. Techniques such as direct coupling analysis can discriminate direct structural contact by measuring the relationship between two positions of a biological sequence

and using statistical models to exclude the effects from other positions (transitive correlations between pairs of residues) (98, 152). Although these approaches have proven very efficient to discriminate direct from transitive three-dimensional contacts in macromolecules, they require large numbers of homologous sequences that are not available for many protein families and may never be for rapidly evolving or recently evolved ones.

Two recent studies have succeeded in distinguishing direct structural contacts using a complete double-mutant landscape generated in a deep mutational scan of protein GB1 (129, 137). One study did so by quantifying interactions directly (129). In the other (137), contacts were predicted using two metrics: the enrichment of epistatic interactions between pairs of residues and the partial correlations of interaction profiles between all pairs of positions within the domain (mutations in positions that are close in space will often interact similarly with the remaining protein residues, unlike mutations in protein positions that are far away from each other; **Figure 5b**). This approach could also predict direct structural contacts from sparse mutagenesis data sets—for example, the RRM2 domain of PAB1 (97) or the WW protein domain of hYAP65 (6), as well as the interface contacts in the protein–protein interaction between FOS and JUN (32).

Epistasis Generated by Changes in Physical Interaction Affinity and Specificity

Many macromolecules function by binding to other molecules. Specific epistasis can appear when mutations disrupt these interactions or promote new promiscuous ones. An example of this type of specific epistasis can be found in proteins binding to different ligands, such as hormones (20, 49, 106), peptides (94, 123), and other proteins (1, 32).

For example, a mutagenesis study of a PDZ domain (PSD95^{pdz3}) found that epistasis between two mutations was sufficient to switch the ligand-binding specificity of PSD95^{pdz3} from its canonical class I ligand toward a class II ligand (94). The first mutation, a type-switching mutation, worked directly and locally in the binding pocket to simultaneously eliminate class I ligand binding and promote class II ligand binding. The second mutation, a type-bridging mutation, worked allosterically to open up the binding pocket, which enabled both class I and class II recognition. The structural rearrangements caused by this second mutation allowed the type-switching mutation to bind to the new ligand, creating a strong positive epistatic signal. To determine whether other mutations in the domain could behave similarly, the authors assayed the binding affinity of all 1,598 single-amino-acid substitutions of the PDZ^{pdz3} domain for the binding to both class I and class II ligands (123). Of the 44 substitutions that changed PDZ^{pdz3} specificity for the class II ligand, 12 of them directly contacted the site of the ligand binding (type-switching mutations), and the remaining 32 were all outside of the contact environment of the new ligand (type-bridging mutations). Almost no positions carried mutations with both roles, suggesting that the distinction between type-switching and type-bridging phenotypes was a characteristic of the position rather than the identity of the mutation.

Another example of specific epistasis arising in physical interactions occurred in the evolution of the specificity in the glucocorticoid receptor for its ligand, the steroid hormone cortisol. Using ancestral reconstruction (**Figure 2c**) and receptor activation assays, Bridgham and colleagues (20, 106) found that the approximately 450-million-year-old precursor of vertebrate glucocorticoid and mineralocorticoid receptors, originally activated by both types of hormone, required a set of historical substitutions to gain cortisol specificity. Of all historical mutations, two substitutions in the ligand-binding pocket (S106P and L111Q) interacted epistatically, inducing a functional switch in the ancestral receptor. Together, both mutations increased the receptor specificity for cortisol over other mineralocorticoids (**Figure 5c**, right). The change of serine 106 to proline repositioned a helix of the ligand-binding pocket, impairing activation by any ligand. However,

this change also repositioned position 111, which allowed the creation of a new cortisol-specific hydrogen bond when mutation L111Q was acquired (106) (**Figure 5c**, left), resulting in an unexpected higher affinity for cortisol (positive epistasis). Nevertheless, the complete switch to cortisol specificity required five additional mutations: three mutations that completed the loss of mineralocorticoid sensitivity but had a destabilizing effect, and two mutations that compensated for the destabilizing effect of the other mutations (106). A later study found that the historical contingency created by these two permissive mutations was highly specific (49). In a library of more than 3,000 variants of the ancestral receptor containing the five corticoid-switch-like mutations, no other permissive mutations could be found apart from the ones introduced by evolution, since the few mutations that rescued the destabilizing effect caused promiscuous activation with other steroids.

Although these previous examples of genetic interactions are highly specific, epistasis involved in the change of affinity for a ligand can often be nonspecific (147). A substitution may have an extremely specific effect on the affinity for a ligand and at the same time alter protein stability. Therefore, an additional mutation is required to compensate for this detrimental effect (15, 16, 151). This principle is illustrated by the gain of resistance of N1 influenza virus to oseltamivir. Mutation H274Y of the viral enzyme neuraminidase confers oseltamivir resistance but compromises the fitness of the virus (130) by decreasing the amount of neuraminidase protein reaching the cell surface. However, viral fitness could be restored when multiple mutations nonspecifically increased the amount of protein that reached the cell surface without interacting specifically with the resistance mutation (15). Similarly, trimethoprim resistance in *E. coli* can be achieved by mutations in the dihydrofolate reductase that impair drug binding. Using experimental evolution, a study showed that such mutations required the prior acquisition of a promoter mutation that increased dihydrofolate reductase expression (151). These cases illustrate that specific and nonspecific epistasis concur, and both play important roles in the specificity for ligand binding.

Interactions Between Proteins and DNA or RNA

Many processes in the cell involve the interaction of proteins with DNA or RNA. For instance, transcription factors change the expression of downstream genes by specifically binding to DNA motifs. Interaction between proteins and DNA-binding motifs can cause specific epistasis when two mutations do not behave additively at the level of free energy of binding (ΔG_B) (4). By contrast, nonspecific interactions occur because of cooperativity between transcription factors (93, 99) or the nonlinear relationship between binding affinity and downstream transcriptional activation (142).

Measuring the binding affinity of transcription factors and DNA-binding sites using high-throughput technologies such as systematic evolution of ligands by exponential enrichment (SELEX) (61) or protein-binding microarrays (102) has revealed abundant epistasis between mutations in DNA-binding motifs (3, 61), which was also shown by deep mutational scans of mutations in protein DNA-binding domains (4, 144). Still, few studies have elucidated the mechanistic causes of such specific interactions (4, 100).

Although most transcription factors bind a single DNA motif, some can bind to different DNA sequences with comparable affinities (61, 161). In the first case, mutations in the transcription-factor-binding motif additively affect the free energy of binding, whereas in the second situation, the effects of DNA substitutions are not independent and instead display a strong epistatic behavior (nonadditive at the level of ΔG_B ; **Figure 5d**). One study disentangled the molecular mechanisms by which the transcription factors HOXB13 and CDX2 could bind two different

Historical contingency: the idea that the paths of evolution are constrained by historical events (i.e., previous mutations)



Diminishing-returns epistasis: epistasis in which the effect on fitness of two mutations is less radical (less beneficial or less detrimental) than expected

DNA motifs with very similar affinities (100). The crystal structures of the transcription factors bound to the two distinct high-affinity sequences revealed no protein structural changes. Thermodynamic analyses showed that the ΔG_B of both complexes was the same in both states; however, the relative contributions of entropy and enthalpy to the free energy of the complexes were dramatically different. In one case, the protein–DNA interaction was assisted by rigid water-mediated hydrogen bonds, increasing enthalpy due to reduction in entropy. By contrast, the alternative DNA–protein complex had much higher entropy because binding involved direct DNA–protein contacts and water moved freely.

Epistasis can also appear when mutations create or destroy RNA-binding motifs and consequently affect RNA localization, degradation, or processing. After accounting for nonspecific epistasis in a deep mutational scan of the alternative exon 6 of *FAS*, Baeza-Centurion et al. (9) found that few specific interactions remained, all occurring between mutations separated by fewer than six nucleotides, within the range of the RNA-binding motifs [four to seven nucleotides (29)], suggestive of epistasis arising due to the binding of proteins. The authors proposed several mechanisms behind these specific epistatic interactions. For instance, positive diminishing-returns epistasis (**Figure 1**) occurs when two mutations independently create two overlapping RNA motifs for the same splicing factor (**Figure 5e**, top). Because the two binding sites overlap, the motifs cannot be bound at the same time, and the double mutation results in a smaller effect than the sum of the individual mutation effects. Alternatively, sign epistasis can emerge when one mutation alone destroys a silencer-binding site (increasing exon inclusion) and the second mutation alone has no effect because it occurs outside of the binding region. Yet when the two mutations combine, they create a new binding site for the same or a new repressor, further decreasing exon inclusion levels (**Figure 5e**, bottom).

HIGHER-ORDER EPISTASIS

Pairwise epistasis can contribute substantially to phenotypic variation between individuals (40). However, interactions between two mutations provide only a limited view of the vast combinatorial size of genotype space (119). Genetic interactions are moderately or poorly conserved among species (33, 128, 149), and mutations with different consequences in different yeast strains are almost always a consequence of complex interactions involving multiple other loci rather than being explained by pairwise interactions (35, 103). These observations point toward the importance of higher-order epistasis, which happens when the interaction between two or more mutations changes in the presence or absence of an additional mutation. For instance, a third-order interaction implies that the pairwise interaction of two mutations is dependent on the presence of a third mutation (**Figure 6a**).

Evidence for higher-order interactions is found in studies ranging from viral to mammalian proteins involving few loci substitutions (132, 155). Although higher-order interactions made small but detectable contributions to these small-scale genotype-to-phenotype maps, more systematic studies were required to decipher their overall importance for evolution and genetic prediction. Sign epistasis between two loci constrains adaptation by limiting the number of selectively accessible paths (154). Do high-order interactions also alter evolutionary outcomes? Are they necessary to predict phenotypes from changes in genotype? Emergent technologies have made it possible to answer these questions by creating all possible combinations of mutations within a limited set of positions in a molecule. For instance, a fitness landscape of four sites in protein GB1, containing $20^4 = 160,000$ variants, found that direct paths for adaptation blocked by pairwise reciprocal sign epistasis can be circumvented by indirect paths involving the gain and subsequent loss of mutations (157) (**Figure 6b**). Similar findings have been reported in fitness landscapes of antibiotic

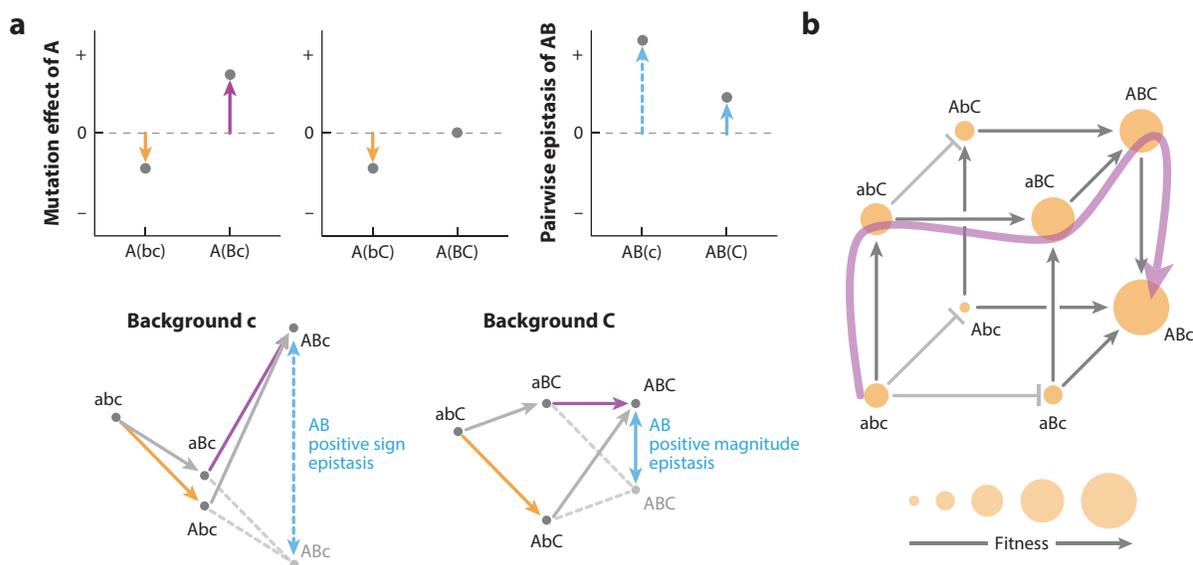


Figure 6

Higher-order genetic interactions. (a) The basis of a third-order genetic interaction. The effect of mutation A (*left*) or the pairwise interactions between mutations A and B (*right*) changes depending on the background. When mutation C is absent (denoted here as “c”), mutation A changes sign with B in the background (positive sign epistasis). However, when mutation C is present (denoted here as “C”), the mutation effect of A increases in magnitude only when it is present together with B (positive magnitude epistasis). The third-order interaction term between mutations A, B, and C corresponds to the difference between the pairwise interaction of mutations A and B when C is present or absent in the genetic background. (b) Higher-order epistasis shaping evolutionary trajectories. Due to the third-order interaction of mutations A, B, and C, direct paths from the wild-type state (*abc*) to the fittest genotype (*ABC*) are not accessible, and additional mutational steps are required.

resistance, where higher-order interactions increase the accessibility to the fittest genotype (111). A recent study generated all mutant combinations of 14 naturally occurring substitutions in a yeast tRNA, showing that all pairs of mutations switched from interacting positively to interacting negatively when found in different genetic backgrounds due to abundant higher-order interactions (34). Although higher-order interactions have been detected within (13, 111, 116, 117) and between (75) genes, most studies have not disentangled their specific and nonspecific components, and a deep understanding of their underlying mechanisms is lacking.

THE ENVIRONMENTAL DEPENDENCE OF EPISTASIS

The fact that mutations can have different effects when the environment changes (12, 51–53, 146) or fluctuates (134) suggests that epistatic interactions can also be environment dependent. Evidence for the dependence of epistasis on the environment has been observed by simulating intracellular growth of phages (159) and modeling metabolic networks in yeast (50). Additionally, gene–gene–environment interactions have been detected in small-scale experiments in viruses (76), bacteria (126), yeast (10, 44), and flies (160). However, these represent a limited view of the possible genotype–environment space. How often do gene–gene–environment interactions occur? Can we predict how genetic interactions vary with the environment? To answer these questions, we need to exhaustively and systematically study epistasis under the influence of variable environments. Nearly all systematic studies of genetic interactions and their implications

at the evolutionary level have been performed in a single static environment, and only recently have mutant libraries been subjected to selection in different environments (85). To what extent the dependence of interactions on the environment is predictable or generalizable across different types of genes is still unknown.

IMPLICATIONS AND APPLICATIONS OF EPISTASIS

Genetic Prediction

A long-standing goal of genetics has been to predict changes in phenotype from genotype changes (80, 83). The definition of epistasis implies that it is difficult to know beforehand what will happen when a mutation occurs. But does epistasis hinder the prediction of phenotypes from genotypes? To what extent are pairwise and higher-order interactions necessary to make accurate predictions? Will prediction be easier if we consider the mechanisms giving rise to epistasis?

In the absence of a mechanistic model, researchers have attempted to fit models allowing for every possible pairwise interaction between mutations (54, 109). However, such models are often overfitted, leading to poor or biased predictions (36, 110). Being aware of nonspecific epistasis and adapting the null model for how mutations combine accordingly (9, 32) reduces the number of interaction terms needed to make accurate genetic predictions. Indeed, in a recent study of an alternative exon (9), the phenotypes (exon inclusion levels) of genotypes containing as many as 10 mutations were predicted using a mechanistic model that included a minimal number of pairwise interactions. Accounting for the global shape of the genotype–phenotype map does not solve all the problems associated with genetic prediction. Even when global trends are taken into account, many significant pairwise and higher-order interactions may remain, hindering the prediction of phenotypes from genotypes (116, 132, 133). Additionally, taking the average effect of mutations across all genotypes in a data set can often lead to accurate predictions even when using a model with few interaction terms (34, 116). To make matters more complicated, the presence of a peaked fitness landscape makes it impossible to estimate the additive biophysical traits of the landscape from phenotype measurements alone, since the same phenotype can be linked to different underlying biophysical states (25, 68, 86) (**Figure 4g**), and a simple environmental change can alter the topology of the local fitness landscape (68).

Molecular Evolution and Infectious Diseases

Because epistasis makes genetic prediction difficult, it also hinders the prediction of evolutionary outcomes, since this prediction relies on knowing which changes in genotype are beneficial or detrimental (13, 30). The contingency of new mutations on previous mutations acquired throughout the course of evolution limits which mutations can be acquired by a population (139, 143) and which evolutionary paths become accessible (157). If two isolated populations of the same species progressively acquire different mutations, the new evolutionary paths available to each population will become progressively more and more different, up to the point where mutations tolerated in one population are deleterious in the other, ultimately leading to speciation (72, 120). Additionally, the sigmoidal relationship between an underlying additive trait and the phenotype seen by natural selection can confer mutational robustness, so that multiple mutations are needed to achieve a phenotypic change if the current gene is at or near the asymptote of the sigmoid (**Figure 4b**). For example, a very stable protein will require multiple destabilizing mutations to reduce its fraction folded (94, 146).

Since epistasis has such profound implications for molecular evolution, it will also have important consequences for health care, because many pathogens are constantly and rapidly evolving

to overcome the immune system and current treatments. One of the main treatments against influenza is the antiviral oseltamivir (101), which blocks the active site of an enzyme essential for viral replication. Influenza strains that carried mutation H274Y and were resistant to oseltamivir were detected during clinical trials, but since H274Y led to attenuated viruses, researchers concluded that this mutation was not clinically relevant (60). Eight years after oseltamivir was introduced into the drug market, antiviral-resistant strains appeared containing mutation H274Y, which spread to most viruses within a year (101). Importantly, this mutation was not deleterious in the newly evolved strains, because before the virus acquired the drug-resistance-conferring mutation, it evolved additional mutations that allowed it to tolerate the originally deleterious mutation (15). Indeed, epistasis has played an important role in shaping the molecular evolution of influenza (74) and HIV (67), as well as in the emergence of antibiotic resistance (111, 130, 154). Understanding how mutations in pathogens interact should make it possible to better predict the course of pathogen evolution and improve the development of vaccines.

Cancer and Autoimmune Diseases

Beyond pathogen evolution, epistasis is also important in cancer. Cancer-causing mutations often interact strongly, with many driver mutations having little effect alone but lethal consequences when combined [sometimes referred to as oncogene cooperation (7)]. Relatively little is understood about why particular driver mutations interact and why this changes across cell types (112). Moreover, exploiting epistasis is now a major strategy for cancer drug development, with large-scale efforts under way to identify gene inhibitions and drug treatments that are synthetic lethal with cancer driver mutations (104). For example, poly (ADP-ribose) polymerase (PARP) inhibitors cause synthetic lethality in tumors carrying mutations in *BRCA2* or in other genes involved in homologous recombination (7). The goal of identifying new synthetic lethal combinations is one of the major motivations for systematically mapping genetic interactions between human genes (95).

Epistatic interactions have also been detected in additional human disease—for example, in autoimmune diseases involving mutations in the human leukocyte antigen (HLA) locus (92). Genes in this region are highly polymorphic, which has enabled studies of how genetic variants combine to influence predisposition to many autoimmune diseases. One such disease is ankylosing spondylitis, an inflammatory arthritis that targets the spine and pelvis of affected individuals. Population studies have revealed that, although more than 90% of ankylosing spondylitis patients carry a specific variant of the *HLA-B* gene (*HLA-B27*), only 5% of *HLA-B27*-positive individuals ever develop the disease (141). This suggests that, in addition to the *HLA-B27* gene, other genetic components may be important for developing the disease. Indeed, the loss of function of *ERAP1*, a non-HLA gene, reduces the risk of ankylosing spondylitis in individuals with the *HLA-B27* variant (27). This interaction is highly informative about the potential mechanism underlying this disease, since the main function of the ERAP1 enzyme is to trim peptides in the endoplasmic reticulum to optimal length such that they can be presented by the HLA-B molecule (8). Similar interactions between *ERAP1* and HLA genes have been found in other human diseases, such as psoriasis (43) and Behçet's disease (71).

Interactions among HLA genes are also common and are often caused by amino acid variation in the antigen-presenting groove of HLA molecules. Mutations in this domain alter the repertoire of peptides presented and increase the risk of autoimmunity, as has been observed in celiac disease (82). The DQ2.5 haplotype, made up of two variants in the antigen-presenting groove of HLA-DQ (HLA-DQA1*05:01 and HLA-DQB1*02:01), is one of the main contributors to celiac disease susceptibility (96), whereas the DQ7 haplotype (made up of alleles also located in the *HLA-DQ*

Synthetic lethality:
a simultaneous perturbation of two genes that results in cellular or organismal death

Human leukocyte antigen (HLA) locus:
a region in the human genome with genes encoding cell-surface proteins that display antigenic peptides to effector immune cells to regulate self-tolerance and downstream immune responses

gene) has not been shown to increase disease risk except in a DQ2.5-positive background (82). Since all the mutations involved are located in the antigen-presenting pocket, these results suggest a mechanism involving the abnormal presentation of gluten by HLA molecules to T cells. Similar interactions among HLA genes have been found in type I diabetes (58).

In other diseases, where the causal antigen is not known, establishing the mechanisms of genetic interactions is more challenging. One example is multiple sclerosis, where the HLA-DRB1*15:01 variant has been strongly associated with the disease. When this variant was found in combination with HLA-DQA1*01:01 (not previously linked to this disease), it conferred a strong protective effect against the disease (59). In this study, however, it was difficult to establish a mechanism underlying the interaction because HLA-DQA1*01:01 is in strong linkage disequilibrium with other genetic variants that might be responsible for the interaction. Unfortunately, this problem is not restricted to this one study: The contribution of epistasis to human disease predisposition is still an open question for most rare and common diseases, and methods to detect interactions suffer from a lack of statistical power (88, 131).

Using Epistasis to Solve Protein Structures and Engineer Proteins

The close relationship between structure and function means that determining the three-dimensional structure of macromolecules is a major goal in biology. Although the past few decades have brought tremendous progress in the experimental determination of three-dimensional protein structures, this process is still cumbersome and costly, driving the search for alternative methods to predict protein structure (91). Since genomic sequences contain evolutionary information about the functional constraints of proteins, recent techniques have taken advantage of the enormous amount of sequence information available. One of these techniques is direct coupling analysis (98, 152), which assumes that coevolving amino acids interact structurally or functionally. Direct coupling analysis has been used to infer residue contacts in known and unknown protein (90, 91) and RNA (153) structures, as well as structural changes due to complex formation or conformational plasticity (55). However, as explained above, this technique may not be feasible for the analysis of fast-evolving, recently evolved, and de novo designed proteins and RNAs.

Since amino acids making direct structural contacts within a protein can be discriminated by their patterns of genetic interactions (see the section titled Specific Epistasis Due to the Three-Dimensional Structure of Molecules), deep mutational scans can provide sufficient information to determine the three-dimensional folds of macromolecules (129, 137). This suggests a new experimental approach to solve macromolecular structures, which requires a selection assay for the activity of the protein or RNA of interest; thus, the adoption or development of generic assays that read out the folding of different molecules is an important challenge. This approach could be particularly useful for studying structures that are difficult to determine using physical techniques, such as disordered and membrane proteins. Moreover, through the use of in vivo selection assays, deep mutational scans have the potential to reveal the in vivo conformations of molecules as they are performing particular functions. A generic approach for in vivo structural biology could have many exciting possibilities for cell biology.

The use of epistatic patterns to predict a particular unknown biophysical trait could be extended to other fields in biology. Since epistasis can reveal how proteins are structured and physically interacting, epistasis analysis may help in protein design. For instance, the unknown structures of de novo designed proteins could be assayed using deep mutational scans to quickly assess whether they are correctly folded. Thus, systematically studying epistasis and its underlying mechanisms not only is important for better understanding genotype–phenotype maps and improving genetic

prediction, but also may have an important impact on fields as diverse as structural biology, synthetic biology, and medicine.

SUMMARY POINTS

1. Epistasis is prevalent within and between genes across many different organisms.
2. Two major types of epistasis are defined depending on whether the interaction depends only on the effect size of the mutations involved (nonspecific epistasis) or also on their identities (specific epistasis).
3. Nonspecific epistasis emerges from a nonlinear relationship between the measured phenotype and the underlying biophysical traits on which mutations have an effect.
4. Mechanisms giving rise to nonspecific epistasis include protein-folding thermodynamics, cooperativity, mutually exclusive molecular competition, and nonlinear expression–fitness functions.
5. Specific epistasis occurs when mutations do not act additively at the level of the underlying biophysical traits.
6. Mechanisms giving rise to specific epistasis include mutations being in close proximity within the three-dimensional structure of a macromolecule, creating new specificities for ligands or partners, and changing the affinity for DNA- or RNA-binding motifs, among others.
7. Pairwise interactions can change in the presence of additional mutations in the genetic background (higher-order epistasis) or due to changes in the environment, with evolutionary implications.
8. Knowledge about epistasis has broad implications for health care and personalized medicine, since it can help to predict how pathogens or a patient will respond to treatments, as well as potential biotechnological applications—for example, to determine the three-dimensional structures of macromolecules.

FUTURE ISSUES

1. Additional mechanisms underlying epistasis need to be investigated to arrive at a more complete understanding of mutation effects.
2. A framework is needed to compare analyses of epistasis across laboratories and data sets.
3. Methods to assess intergenic and intragenic epistasis in the same assay will need to be developed to better understand their relative importance and how a mutation's effects propagate across different layers of biological organization.
4. The role of the environment in altering epistatic interactions is currently unclear, and this will need to be addressed systematically in future studies.
5. Systematic analysis of epistasis between disease-causing mutations (e.g., driver cancer mutations) will be necessary to provide us with a better understanding of the molecular bases of disease and their possible treatments.
6. Deep mutagenesis of diverse macromolecules is required to better understand how best to use genetic interactions to predict three-dimensional structures.



DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by the European Research Council (Consolidator Grant 616434), the Spanish Ministry of Economy and Competitiveness (grants BFU2017-89488-P and SEV-2012-0208), the AXA Research Fund, the Bettencourt Schueller Foundation, Agència de Gestió d'Ajuts Universitaris i de Recerca (grant SGR-831), and the Centres de Recerca de Catalunya Program/Generalitat de Catalunya. P.B.-C. was funded in part by a Severo Ochoa PhD fellowship. We thank all the members of the Lehner laboratory, especially Guillaume Diss and Jörn Schmedel.

LITERATURE CITED

1. Aakre CD, Herrou J, Phung TN, Perchuk BS, Crosson S, Laub MT. 2015. Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* 163:594–606
2. Ackers GK, Johnson AD, Shea MA. 1982. Quantitative model for gene regulation by λ phage repressor. *PNAS* 79:1129–33
3. Aguilar-Rodríguez J, Payne JL, Wagner A. 2017. A thousand empirical adaptive landscapes and their navigability. *Nat. Ecol. Evol.* 1:0045
4. Anderson DW, McKeown AN, Thornton JW. 2015. Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *eLife* 4:e07864
5. Araya CL, Fowler DM. 2011. Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.* 29:435–42
6. Araya CL, Fowler DM, Chen W, Muniez I, Kelly JW, Fields S. 2012. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *PNAS* 109:16858–63
7. Ashworth A, Lord CJ, Reis-Filho JS. 2011. Genetic interactions in cancer progression and treatment. *Cell* 145:30–38
8. Australo-Anglo-Am. Spondyloarthritis Consort., Wellcome Trust Case Control Consort. 2, Evans DM, Spencer CCA, Pointon JJ, et al. 2011. Interaction between *ERAPI* and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat. Genet.* 43:761–67
9. Baeza-Centurion P, Miñana B, Schiedel JM, Valcárcel J, Lehner B. 2019. Combinatorial genetics reveals a scaling law for the effects of mutations on splicing. *Cell* 176:549–63
10. Bandyopadhyay S, Mehta M, Kuo D, Sung M-K, Chuang R, et al. 2010. Rewiring of genetic networks in response to DNA damage. *Science* 330:1385–89
11. Bank C, Hietpas RT, Jensen JD, Bolon DNA. 2015. A systematic survey of an intragenic epistatic landscape. *Mol. Biol. Evol.* 32:229–38
12. Bank C, Hietpas RT, Wong A, Bolon DN, Jensen JD. 2014. A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. *Genetics* 196:841–52
13. Bank C, Matuszewski S, Hietpas RT, Jensen JD. 2016. On the (un)predictability of a large intragenic fitness landscape. *PNAS* 113:14085–90
14. Bateson. 1909. Discussion on the influence of heredity on disease, with special reference to tuberculosis, cancer, and diseases of the nervous system: introductory address. *Proc. R. Soc. Med.* 2(Gen. Rep.):22–30
15. Bloom JD, Gong LI, Baltimore D. 2010. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* 328:1272–75
16. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. 2006. Protein stability promotes evolvability. *PNAS* 103:5869–74

17.22 Domingo • Baeza-Centurion • Lehner

Review in Advance first posted on
May 13, 2019. (Changes may still
occur before final publication.)



17. Boj SF, Petrov D, Ferrer J. 2010. Epistasis of transcriptomes reveals synergism between transcriptional activators Hnf1 α and Hnf4 α . *PLoS Genet.* 6:e1000970
18. Braun S, Enculescu M, Setty ST, Cortés-López M, de Almeida BP, et al. 2018. Decoding a cancer-relevant splicing decision in the *RON* proto-oncogene using high-throughput mutagenesis. *Nat. Commun.* 9:3315
19. Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA. 2012. Epistasis as the primary factor in molecular evolution. *Nature* 490:535–38
20. Bridgham JT, Carroll SM, Thornton JW. 2006. Evolution of hormone-receptor complexity by molecular exploitation. *Science* 312:97–101
21. Burga A, Casanueva MO, Lehner B. 2011. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature* 480:250–53
22. Burga A, Lehner B. 2012. Beyond genotype to phenotype: why the phenotype of an individual cannot always be predicted from their genome sequence and the environment that they experience. *FEBS J.* 279:3765–75
23. Burga A, Lehner B. 2013. Predicting phenotypic variation from genotypes, phenotypes and a combination of the two. *Curr. Opin. Biotechnol.* 24:803–9
24. Casanueva MO, Burga A, Lehner B. 2012. Fitness trade-offs and environmentally induced mutation buffering in isogenic *C. elegans*. *Science* 335:82–85
25. Chou H-H, Delaney NF, Draghi JA, Marx CJ. 2014. Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLoS Genet.* 10:e1004149
26. Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF. 2013. Genetic incompatibilities are widespread within species. *Nature* 504:135–37
27. Cortes A, Pulit SL, Leo PJ, Pointon JJ, Robinson PC, et al. Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with *ERAP1*. *Nat. Commun.* 6:7146
28. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, et al. 2016. A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353:aaf1420
29. Daubner GM, Cléry A, Allain FH-T. 2013. RRM-RNA recognition: NMR or crystallography...and new findings. *Curr. Opin. Struct. Biol.* 23:100–8
30. de Visser JAGM, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* 15:480–90
31. Dekel E, Alon U. 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436:588–92
32. Diss G, Lehner B. 2018. The genetic landscape of a physical interaction. *eLife* 7:e32472
33. Dixon SJ, Andrews BJ, Boone C. 2009. Exploring the conservation of synthetic lethal genetic interaction networks. *Commun. Integr. Biol.* 2:78–81
34. Domingo J, Diss G, Lehner B. 2018. Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature* 558:117–21
35. Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, et al. 2010. Genotype to phenotype: a complex problem. *Science* 328:469
36. du Plessis L, Leventhal GE, Bonhoeffer S. 2016. How good are statistical models at approximating complex fitness landscapes? *Mol. Biol. Evol.* 33:2454–68
37. Eldar A, Elowitz MB. 2010. Functional roles for noise in genetic circuits. *Nature* 467:167–73
38. Fischer B, Sandmann T, Horn T, Billmann M, Chaudhary V, et al. 2015. A map of directional genetic interactions in a metazoan cell. *eLife* 4:e05464
39. Fisher RA. 1919. XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52:399–433
40. Forsberg SKG, Bloom JS, Sadhu MJ, Kruglyak L, Ö Carlborg. 2017. Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat. Genet.* 49:497–503
41. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, et al. 2010. High-resolution mapping of protein sequence-function relationships. *Nat. Methods* 7:741–46



42. Gao L, Zhang J. 2003. Why are some human disease-associated mutations fixed in mice? *Trends Genet.* 19:678–81
43. Genet. Anal. Psoriasis Consort., Wellcome Trust Case Control Consort. 2. 2010. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*. *Nat. Genet.* 42:985–90
44. Gerke J, Lorenz K, Ramnarine S, Cohen B. 2010. Gene-environment interactions at nucleotide resolution. *PLoS Genet.* 6:e1001144
45. Goodwin EB, Ellis RE. 2002. Turning clustering loops: sex determination in *Caenorhabditis elegans*. *Curr. Biol.* 12:R111–20
46. Guy MP, Young DL, Payea MJ, Zhang X, Kon Y, et al. 2014. Identification of the determinants of tRNA function and susceptibility to rapid tRNA decay by high-throughput in vivo analysis. *Genes Dev.* 28:1721–32
47. Haerd T, Dahlman K, Carlstedt-Duke J, Gustafsson JA, Rigler R. 1990. Cooperativity and specificity in the interactions between DNA and the glucocorticoid receptor DNA binding domain. *Biochemistry* 29:5358–64
48. Halabi N, Rivoire O, Leibler S, Ranganathan R. 2009. Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138:774–86
49. Harms MJ, Thornton JW. 2014. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* 512:203–7
50. Harrison R, Papp B, Pál C, Oliver SG, Delneri D. 2007. Plasticity of genetic interactions in metabolic networks of yeast. *PNAS* 104:2307–12
51. Hayden EJ, Bendixsen DP, Wagner A. 2015. Intramolecular phenotypic capacitance in a modular RNA molecule. *PNAS* 112:12444–49
52. Hayden EJ, Ferrada E, Wagner A. 2011. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474:92–95
53. Hietpas RT, Bank C, Jensen JD, Bolon DNA. 2013. Shifting fitness landscapes in response to altered environments. *Evolution* 67:3512–22
54. Hinkley T, Martins J, Chappay C, Haddad M, Stawiski E, et al. 2011. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nat. Genet.* 43:487–89
55. Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS. 2012. Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* 149:1607–21
56. Horlbeck MA, Xu A, Wang M, Bennett NK, Park CY, et al. 2018. Mapping the genetic landscape of human cells. *Cell* 174:953–67.e22
57. Horovitz A. 1996. Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold. Des.* 1:R121–26
58. Hu X, Deutsch AJ, Lenz TL, Onengut-Gumuscu S, Han B, et al. 2015. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat. Genet.* 47:898–905
59. Int. Mult. Scler. Genet. Consort. 2015. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat. Genet.* 47:1107–13
60. Ives JAL, Carr JA, Mendel DB, Tai CY, Lambkin R, et al. 2002. The H274Y mutation in the influenza A/H1N1 neuraminidase active site following oseltamivir phosphate treatment leave virus severely compromised both in vitro and in vivo. *Antiviral Res.* 55:307–17
61. Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, et al. 2013. DNA-binding specificities of human transcription factors. *Cell* 152:327–39
62. Jordan DM, Frangakis SG, Golzio C, Cassa CA, Kurtzberg J, et al. 2015. Identification of *cis*-suppression of human disease mutations by comparative genomics. *Nature* 524:225–29
63. Julien P, Miñana B, Baeza-Centurion P, Valcárcel J, Lehner B. 2016. The complete local genotype-phenotype landscape for the alternative splicing of a human exon. *Nat. Commun.* 7:11558
64. Kacser H, Burns JA. 1981. The molecular basis of dominance. *Genetics* 97:639–66
65. Kammenga JE. 2017. The background puzzle: how identical mutations in the same gene lead to different disease symptoms. *FEBS J.* 284:3362–73

17.24 Domingo • Baeza-Centurion • Lehner

Review in Advance first posted on
May 13, 2019. (Changes may still
occur before final publication.)



66. Ke S, Shang S, Kalachikov SM, Morozova I, Yu L, et al. 2011. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res.* 21:1360–74
67. Kelleher AD, Long C, Holmes EC, Allen RL, Wilson J, et al. 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *J. Exp. Med.* 193:375–86
68. Kemble HE, Eisenhauer C, Couce A, Chapron A, Magnan M, et al. 2018. Flux, toxicity and protein expression costs shape genetic interaction in a metabolic pathway. bioRxiv 362327. <https://doi.org/10.1101/362327>
69. Keren L, Haussler J, Lotan-Pompan M, Vainberg Slutskin I, Alisar H, et al. 2016. Massively parallel interrogation of the effects of gene expression levels on fitness. *Cell* 166:1282–94.e18
70. Kiel C, Verschuere E, Yang J-S, Serrano L. 2013. Integration of protein abundance and structure data reveals competition in the ErbB signaling network. *Sci. Signal.* 6:ra109
71. Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I. 2013. Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between *HLA-B*51* and *ERAP1*. *Nat. Genet.* 45:202–7
72. Kondrashov AS, Sunyaev S, Kondrashov FA. 2002. Dobzhansky-Muller incompatibilities in protein evolution. *PNAS* 99:14878–83
73. Kondrashov FA, Kondrashov AS. 2001. Multidimensional epistasis and the disadvantage of sex. *PNAS* 98:12089–92
74. Kryazhimskiy S, Dushoff J, Bazykin GA, Plotkin JB. 2011. Prevalence of epistasis in the evolution of influenza A surface proteins. *PLOS Genet.* 7:e1001301
75. Kuzmin E, VanderSluis B, Wang W, Tan G, Deshpande R, et al. 2018. Systematic analysis of complex genetic interactions. *Science* 360:eaa01729
76. Lalić J, Elena SF. 2013. Epistasis between mutations is host-dependent for an RNA virus. *Biol. Lett.* 9:20120396
77. Laufer C, Fischer B, Billmann M, Huber W, Boutros M. 2013. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat. Methods* 10:427–31
78. Laufer C, Fischer B, Huber W, Boutros M. 2014. Measuring genetic interactions in human cells by RNAi and imaging. *Nat. Protoc.* 9:2341–53
79. Lehner B. 2011. Molecular mechanisms of epistasis within and between genes. *Trends Genet.* 27:323–31
80. Lehner B. 2013. Genotype to phenotype: lessons from model organisms for human genetics. *Nat. Rev. Genet.* 14:168–78
81. Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. 2006. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat. Genet.* 38:896–903
82. Lenz TL, Deutsch AJ, Han B, Hu X, Okada Y, et al. 2015. Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. *Nat. Genet.* 47:1085–90
83. Lewontin RC. 1974. *The Genetic Basis of Evolutionary Change*. New York: Columbia Univ. Press
84. Li C, Qian W, Maclean CJ, Zhang J. 2016. The fitness landscape of a tRNA gene. *Science* 352:837–40
85. Li C, Zhang J. 2018. Multi-environment fitness landscapes of a tRNA gene. *Nat. Ecol. Evol.* 2:1025–32
86. Li X, Lalic C, Baeza-Centurion P, Dhar R, Lehner B. 2019. Changes in gene expression shift and switch genetic interactions. bioRxiv 578419. <https://doi.org/10.1101/578419>
87. Lunzer M, Miller SP, Felsheim R, Dean AM. 2005. The biochemical architecture of an ancient adaptive landscape. *Science* 310:499–501
88. Mackay TFC, Moore JH. 2014. Why epistasis is important for tackling complex human disease genetics. *Genome Med.* 6:42
89. Mani R, St. Onge RP, Hartman JL IV, Giaever G, Roth FP. 2008. Defining genetic interaction. *PNAS* 105:3461–66
90. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, et al. 2011. Protein 3D structure computed from evolutionary sequence variation. *PLOS ONE* 6:e28766
91. Marks DS, Hopf TA, Sander C. 2012. Protein structure prediction from sequence variation. *Nat. Biotechnol.* 30:1072–80



92. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. 2017. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol.* 18:76
93. McKeown AN, Bridgham JT, Anderson DW, Murphy MN, Ortlund EA, Thornton JW. 2014. Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell* 159:58–68
94. McLaughlin RN Jr., Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. 2012. The spatial architecture of protein function and adaptation. *Nature* 491:138–42
95. McLornan DP, List A, Mufti GJ. 2014. Applying synthetic lethality for the selective targeting of cancer. *N. Engl. J. Med.* 371:1725–35
96. Megiorni F, Pizzuti A. 2012. *HLA-DQA1* and *HLA-DQB1* in celiac disease predisposition: practical implications of the HLA molecular typing. *J. Biomed. Sci.* 19:88
97. Melamed D, Young DL, Gamble CE, Miller CR, Fields S. 2013. Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein. *RNA* 19:1537–51
98. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, et al. 2011. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *PNAS* 108:E1293–301
99. Morgunova E, Taipale J. 2017. Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.* 47:1–8
100. Morgunova E, Yin Y, Das PK, Jolma A, Zhu F, et al. 2018. Two distinct DNA sequences recognized by transcription factors represent enthalpy and entropy optima. *eLife* 7:e32963
101. Moscona A. 2009. Global transmission of oseltamivir-resistant influenza. *N. Engl. J. Med.* 360:953–56
102. Mukherjee S, Berger MF, Jona G, Wang XS, Muzzey D, et al. 2004. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.* 36:1331–39
103. Mullis MN, Matsui T, Schell R, Force R, Ehrenreich IM. 2018. The complex underpinnings of genetic background effects. *Nat. Commun.* 9:3548
104. Nijman SMB, Friend SH. 2013. Potential of the synthetic lethality principle. *Science* 342:809–11
105. Olson CA, Wu NC, Sun R. 2014. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* 24:2643–51
106. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317:1544–48
107. Otwinowski J. 2018. Biophysical inference of epistasis and the effects of mutations on protein stability and function. *Mol. Biol. Evol.* 35:2345–54
108. Otwinowski J, McCandlish DM, Plotkin JB. 2018. Inferring the shape of global epistasis. *PNAS* 115:E7550–58
109. Otwinowski J, Nemenman I. 2013. Genotype to phenotype mapping and the fitness landscape of the *E. coli lac* promoter. *PLOS ONE* 8:e61570
110. Otwinowski J, Plotkin JB. 2014. Inferring fitness landscapes by regression produces biased estimates of epistasis. *PNAS* 111:E2301–9
111. Palmer AC, Toprak E, Baym M, Kim S, Veres A, et al. 2015. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat. Commun.* 6:7385
112. Park S, Lehner B. 2015. Cancer type-dependent genetic interactions between cancer driver alterations indicate plasticity of epistasis across cell types. *Mol. Syst. Biol.* 11:824
113. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, et al. 2012. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* 30:265–70
114. Phillips PC. 2008. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9:855–67
115. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. 2007. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445:383–86
116. Poelwijk FJ, Socolich M, Ranganathan R. 2017. Learning the pattern of epistasis linking genotype and phenotype in a protein. bioRxiv 213835. <https://doi.org/10.1101/213835>
117. Pokusaeva V, Usmanova D, Putintseva E, Espinar L, Sarkisyan K, et al. 2017. Experimental assay of a fitness landscape on a macroevolutionary scale. bioRxiv 222778. <https://doi.org/10.1101/222778>

118. Poon A, Chao L. 2005. The rate of compensatory mutation in the DNA bacteriophage ϕ X174. *Genetics* 170:989–99
119. Povolotskaya IS, Kondrashov FA. 2010. Sequence space and the ongoing expansion of the protein universe. *Nature* 465:922–26
120. Presgraves DC. 2007. Speciation genetics: epistasis, conflict and the origin of species. *Curr. Biol.* 17:R125–27
121. Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016. Network of epistatic interactions within a yeast snoRNA. *Science* 352:840–44
122. Raj A, van Oudenaarden A. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135:216–26
123. Raman AS, White KI, Ranganathan R. 2016. Origins of allostery and evolvability in proteins: a case study. *Cell* 166:468–80
124. Ranganathan S. 1999. A technique for prediction of carbonitride precipitation in high strength low alloy steels. *Mater. Sci. Technol.* 15:523–26
125. Rauscher B, Heigwer F, Henkel L, Hielscher T, Voloshanenko O, Boutros M. 2018. Toward an integrated map of genetic interactions in cancer cells. *Mol. Syst. Biol.* 14:e7656
126. Remold SK, Lenski RE. 2004. Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. *Nat. Genet.* 36:423–26
127. Rich MS, Payen C, Rubin AF, Ong GT, Sanchez MR, et al. 2016. Comprehensive analysis of the *SUL1* promoter of *Saccharomyces cerevisiae*. *Genetics* 203:191–202
128. Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, et al. 2008. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* 322:405–10
129. Rollins NJ, Brock KP, Poelwijk FJ, Stiffler MA, Gauthier NP, et al. 2018. 3D protein structure from genetic epistasis experiments. bioRxiv 320721. <https://doi.org/10.1101/320721>
130. Russell RJ, Haire LF, Stevens DJ, Collins PJ, Lin YP, et al. 2006. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* 443:45–49
131. Sackton TB, Hartl DL. 2016. Genotypic context and epistasis in individuals and populations. *Cell* 166:279–87
132. Sailer ZR, Harms MJ. 2017. Detecting high-order epistasis in nonlinear genotype-phenotype maps. *Genetics* 205:1079–88
133. Sailer ZR, Harms MJ. 2017. High-order epistasis shapes evolutionary trajectories. *PLoS Comput. Biol.* 13:e1005541
134. Salignon J, Richard M, Fulcrand E, Duplus-Bottin H, Yvert G. 2018. Genomics of cellular proliferation in periodic environmental fluctuations. *Mol. Syst. Biol.* 14:e7823
135. Salinas VH, Ranganathan R. 2018. Coevolution-based inference of amino acid interactions underlying protein function. *eLife* 7:e34300
136. Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, et al. 2016. Local fitness landscape of the green fluorescent protein. *Nature* 533:397–401
137. Schmiedel J, Lehner B. 2018. Determining protein structures using genetics. bioRxiv 303875. <https://doi.org/10.1101/303875>
138. Segrè D, Deluna A, Church GM, Kishony R. 2005. Modular epistasis in yeast metabolism. *Nat. Genet.* 37:77–83
139. Shah P, McCandlish DM, Plotkin JB. 2015. Contingency and entrenchment in protein evolution under purifying selection. *PNAS* 112:E3226–35
140. Shendure J, Akey JM. 2015. The origins, determinants, and consequences of human mutations. *Science* 349:1478–83
141. Sieper J, Rudwaleit M, Khan MA, Braun J. 2006. Concepts and epidemiology of spondyloarthritis. *Best Pract. Res. Clin. Rheumatol.* 20:401–17
142. Sorrells TR, Booth LN, Tuch BB, Johnson AD. 2015. Intersecting transcription networks constrain gene regulatory evolution. *Nature* 523:361–65
143. Starr TN, Flynn JM, Mishra P, Bolon DNA, Thornton JW. 2018. Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *PNAS* 115:4453–58



144. Starr TN, Picton LK, Thornton JW. 2017. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* 549:409–13
145. Sternberg PW, Robert Horvitz H. 1989. The combined action of two intercellular signaling pathways specifies three cell fates during vulval induction in *C. elegans*. *Cell* 58:679–93
146. Stiffler MA, Hekstra DR, Ranganathan R. 2015. Evolvability as a function of purifying selection in TEM-1 β -lactamase. *Cell* 160:882–92
147. Storz JF. 2018. Compensatory mutations and epistasis for protein function. *Curr. Opin. Struct. Biol.* 50:18–25
148. Thomas JH, Birnby DA, Vowels JJ. 1993. Evidence for parallel processing of sensory information controlling dauer formation in *Caenorhabditis elegans*. *Genetics* 134:1105–17
149. Tischler J, Lehner B, Fraser AG. 2008. Evolutionary plasticity of genetic interaction networks. *Nat. Genet.* 40:390–91
150. Tokuriki N, Tawfik DS. 2009. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* 19:596–604
151. Toprak E, Veres A, Michel J-B, Chait R, Hartl DL, Kishony R. 2011. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* 44:101–5
152. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. 2009. Identification of direct residue contacts in protein-protein interaction by message passing. *PNAS* 106:67–72
153. Weinreb C, Riesselman AJ, Ingraham JB, Gross T, Sander C, Marks DS. 2016. 3D RNA and functional interactions from evolutionary couplings. *Cell* 165:963–75
154. Weinreich DM. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111–14
155. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB. 2013. Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* 23:700–7
156. Weinreich DM, Watson RA, Chao L. 2005. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59:1165–74
157. Wu NC, Dai L, Anders Olson C, Lloyd-Smith JO, Sun R. 2016. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* 5:e16965
158. Wylie CS, Shakhnovich EI. 2011. A biophysical protein folding model accounts for most mutational fitness effects in viruses. *PNAS* 108:9916–21
159. You L, Yin J. 2002. Dependence of epistasis on environment and mutation severity as revealed by in silico mutagenesis of phage τ 7. *Genetics* 160:1273–81
160. Zhu C-T, Ingelmo P, Rand DM. 2014. G \times G \times E for lifespan in *Drosophila*: mitochondrial, nuclear, and dietary interactions that modify longevity. *PLoS Genet.* 10:e1004354
161. Zuo Z, Roy B, Chang YK, Granas D, Stormo GD. 2017. Measuring quantitative effects of methylation on transcription factor-DNA binding affinity. *Sci Adv.* 3:eao1799