REVIEWS AND SYNTHESES

Recent insights into the genotype–phenotype relationship from massively parallel genetic assays

Harry Kemble[1,2] iD | Philippe Nghe[2] | Olivier Tenaillon[1] iD

[1]Infection, Antimicrobials, Modelling, Evolution, INSERM, Unité Mixte de Recherche 1137, Université Paris Diderot, Université Paris Nord, Paris, France

[2]École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI Paris), UMR CNRS-ESPCI CBI 8231, PSL Research University, Paris Cedex 05, France

Correspondence
Harry Kemble, Infection, Antimicrobials, Modelling, Evolution, INSERM, Unité Mixte de Recherche 1137, Université Paris Diderot, Université Paris Nord, 75018 Paris, France.
Email: kembleharry@gmail.com

Funding information
FP7 Ideas: European Research Council, Grant/Award Number: 310944

Abstract

With the molecular revolution in Biology, a mechanistic understanding of the genotype–phenotype relationship became possible. Recently, advances in DNA synthesis and sequencing have enabled the development of deep mutational scanning assays, capable of scoring comprehensive libraries of genotypes for fitness and a variety of phenotypes in massively parallel fashion. The resulting empirical genotype–fitness maps pave the way to predictive models, potentially accelerating our ability to anticipate the behaviour of pathogen and cancerous cell populations from sequencing data. Besides from cellular fitness, phenotypes of direct application in industry (*e.g.* enzyme activity) and medicine (*e.g.* antibody binding) can be quantified and even selected directly by these assays. This review discusses the technological basis of and recent developments in massively parallel genetics, along with the trends it is uncovering in the genotype–phenotype relationship (distribution of mutation effects, epistasis), their possible mechanistic bases and future directions for advancing towards the goal of predictive genetics.

KEYWORDS

distribution of fitness effects, epistasis, fitness landscapes, genotype–phenotype maps, high-throughput genetics, phenotypic models

1 | INTRODUCTION

The relationship between genotype, understood in the age of molecular biology as nucleic acid or polypeptide sequence, and phenotype, any trait from the molecular to organismal scale, is of profound importance across biological disciplines. Direct genotype–phenotype mapping has recently become possible on a large scale thanks primarily to the creative use of next-generation sequencing (NGS) technologies to develop massively parallel assays capable of measuring a particular phenotype for each genetic variant in a large mixed library.

One of the most amenable phenotypes to measurement by such assays is microbial or cell fitness, which simply requires the tracking of genotype frequencies during laboratory propagation of a mixed culture (Hietpas, Jensen, & Bolon, 2011). The increasing availability of the resulting genotype–fitness maps opens the door to predictive models, potentially accelerating our ability to anticipate the behaviour of pathogens and cancerous cells from sequencing data. The quantification of other phenotypes alongside fitness, such as expression level or enzymatic activity, can enhance genotype–fitness models by providing mechanistic descriptions of fitness and intermediate phenotypes, themselves of interest to bioengineering (*e.g.* rational design of high-performance enzymes) and precision medicine (Lehner, 2013). Further, the assays presented in the following section can be used directly for the selection of industrially or medically useful phenotypes, as exemplified by the 2018 Nobel Prizes in Chemistry for antibody and enzyme engineering.

This review synthesizes technological developments in massively parallel genetics, the trends it is uncovering in the genotype–phenotype relationship and their potential mechanistic bases. The three main sections review mutation effects in, respectively, single genes, two or a few genes, and entire genomes, with each section discussing the distributions of mutational effects (DME) and the interaction between mutations (epistasis).

2 | ONE GENE

2.1 | Experimental history of large-scale genotype–phenotype mapping for single genes

Empirical snapshots of the local, or occasionally global, genotype–phenotype map for short sequences have recently become possible through combinations of high-throughput technologies. The technical challenge may be broken down as follows: (a) the number of possible nucleic acid or protein sequence variants grows exponentially with sequence length; (b) to draw statistical conclusions about the genotype–phenotype relationship for even very short sequences (*e.g.*>5-mer), a large number (>10^3) of sequence variants must be characterized; (c) currently, the only economical way of generating large numbers of sequence variants is in bulk; and (d) to characterize a phenotype of a large number of pooled sequence variants, not only do we need a high-throughput phenotyping technology, but also a high-throughput way to trace phenotypes back to their sequence.

The earliest strategy used to achieve this last challenge was *phage display* (Scott & Smith, 1990; Smith, 1985). It allows biochemical phenotypes of polypeptides, such as binding affinity, to be easily traced back to their coding DNA sequence by fusing the polypeptide of interest to a viral coat protein and thus making it accessible for biochemical analysis while remaining physically associated with the DNA from which it is expressed. Soon after the invention of phage display, alternative display systems were developed (bacteria, yeast, ribosome, mRNA) (Levin & Weiss, 2006), but it was not until about 20 years after that a display strategy was used to attempt comprehensive local genotype–phenotype characterization (Pál, Kouadio, Artis, Kossiakoff, & Sidhu, 2006).
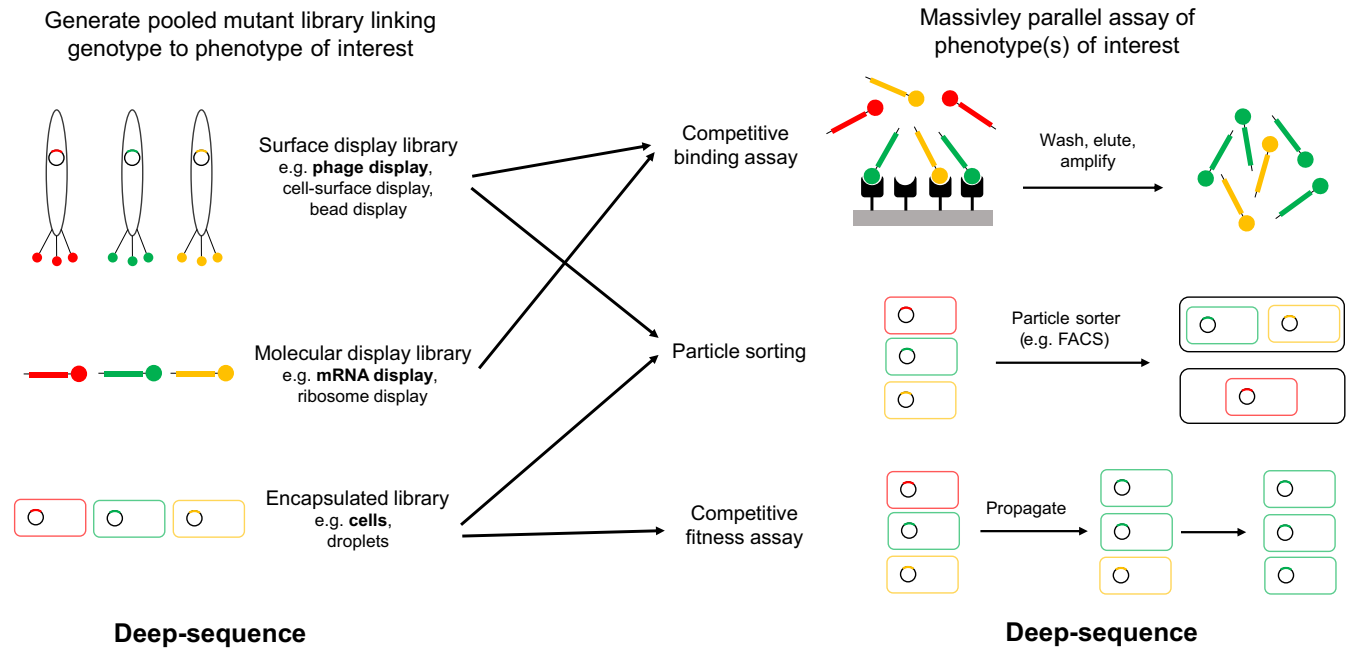
In this study, single point mutations were systematically introduced into the 35-aa receptor-binding site of human growth hormone using Kunkel mutagenesis (Kunkel, Bebenek, & McClary, 1991), a phage display library was constructed, the library was screened for binding to either a structure-specific antibody or the human growth hormone receptor (to assess structural stability and receptor-binding affinity, respectively), and a sample of screened clones was Sanger-sequenced to achieve quasi-quantitative measures for these two phenotypes. This first comprehensive *deep mutational scanning* study produced several novel insights: (a) the native protein fold was highly tolerant to mutations in the targeted solvent-exposed positions (with the exceptions of cysteine and pro-line), with all positions showing a similar level of robustness; (b) hydrophobic residues were generally more stabilizing than hydrophilic ones, a very surprising result for a solvent-exposed region; (c) at the

majority of positions, mutations existed that resulted in both greater stability and stronger receptor binding than the wild type; (d) binding affinity was less robust to mutation than stability, with robustness being position-specific; (e) binding affinity robustness did not relate strongly to sequence conservation across species; and (f) physico-chemical similarity of residue side chains did not correlate strongly with phenotypic effects (Pál et al., 2006).

Although hugely informative, only a few studies employing this methodological strategy have been performed, as it still involves one low-throughput step—Sanger sequencing. Microarray binding assays provide a higher-throughput approach for the specific case of assessing binding of a predefined array of short nucleic acid sequences to a protein (Badis et al., 2009), but ultimately it was the arrival of massively parallel sequencing technologies that allowed large-scale genotype–phenotype mapping studies to flourish. Similarly to the aforementioned Sanger-sequencing study, the earliest such study used phage display of human WW domain variants coupled with (weak) selection for binding to its peptide ligand, with Illumina sequencing of pre- and postselection libraries to again give a quasi-quantitative measure of binding affinity (Fowler et al., 2010). The conclusions differed substantially: 97% of the library variants bound the ligand less tightly than did the wild type; mutational intolerance of the different positions correlated strongly with evolutionary conservation; the core ligand binding region was generally intolerant to mutation; each position appeared to possess a unique mutational preference spectrum; and global thermodynamic stability was concluded to be the primary determinant of binding affinity. Some of the inconsistencies between these two studies may have a technical basis, but they also provide a first hint that the genotype–phenotype relationship, even for a given biochemical phenotype, may vary substantially between proteins.

With the advent of high-throughput sequencing, the limitation becomes the coupling of phenotype measurement with sequencing. Creative approaches have been developed based on surface display to assess a mutant protein library for binding to other proteins, DNA, RNA and small molecule ligands, as well as amenable enzymatic activities such as ubiquitination (Fowler & Fields, 2014; Starita et al., 2013). Further, particle sorting techniques like cell and microflu-idic-droplet sorting can be used to place variants into phenotypic (*e.g.* fluorescence, cell size) bins which are then each subjected to deep-sequencing (Fowler & Fields, 2014; Kinney, Murugan, Callan, & Cox, 2010; McLaughlin, Poelwijk, Raman, Gosal, & Ranganathan, 2012; Noderer et al., 2014; Sarkisyan et al., 2016; Whitehead et al., 2012) (Figure 1).

To overcome this limitation, in vitro fluorescence-coupled microarray and flow-cell techniques have been developed to *directly* measure the biochemical phenotypes of a large number of nucleic acid or protein mutants in parallel, and associate these measurements with the respective genotypes (Boyle et al., 2017; Buenrostro et al., 2014; Geertz, Shore, & Maerkl, 2012; Guenther et al., 2013; Layton, McMahon, & Greenleaf, 2019; Maerkl & Quake, 2009; Nutiu et al., 2011; Shultzaberger, Maerkl, Kirsch, & Eisen, 2012). The current limit of such methods is their requirement for highly specialized

Generate pooled mutant library linking genotype to phenotype of interest

Massivley parallel assay of phenotype(s) of interest

FIGURE 1 Mutant library types and massively parallel sequencing-resolved assays for high-throughput genotype–phenotype mapping. Bulk mutagenesis is used to construct an in vitro or in vivo genotype library, with the phenotype(s) of interest associated with the genotype either by "display" or encapsulation. Phenotypic measurements are then linked to genotypes by deep-sequencing of the library before and after some selection procedure. Selection procedures include binding to a target for direct binding phenotypes, particle sorting for any optically assayable phenotype (*e.g.* fluorescence or cell dimensions) and simple propagation under selective conditions for competitive fitness. Filled rectangles (straight or curved): nucleic acids; filled circles: proteins. Illustrated library examples are named in bold

equipment and expertise, restricting their use to just a handful of laboratories.

One particular phenotype has proven generally amenable to easy and affordable experimental genotype–phenotype coupling, however. It allows precise, quantitative estimates and is of direct evolutionary significance: competitive cellular fitness (Hietpas et al., 2011) (Figure 1). In the original EMPIRIC (*Extremely Methodical and Parallel Investigation of Randomized Individual Codons*) experiment, a 9-a.a. region of *S. cerevisiae* Hsp90 (an essential and highly conserved eukaryotic chaperone protein) was comprehensively mutagenized to create a library of all possible single point mutants. The fitness impact of all mutations was then directly measured in bulk, by serially propagating the library in conditions where expression of the native *hsp90* copy was repressed, and using deep-sequencing to sample mutant frequencies at several time points over a few days. Because the wild-type sequence was included in the library and the wild-type generation time was known, the competitive fitness of each mutant could be estimated relative to the wild type as a selection coefficient, by taking the change in the ratio of mutant to wild-type reads as a function of wild-type generation time.

This data set provided an important empirical insight: the distribution of fitness effects in this individual gene was bimodal, with a fairly equal proportion of mutations being either strongly deleterious or nearly neutral, consistent with a *nearly neutral model* of molecular evolution (Ohta, 1973). Two expected results were also confirmed: synonymous substitutions had far smaller effects on fitness than did nonsynonymous ones (but see also Agashe et al., 2016;

Cuevas, Domingo-Calap, & Sanjuán, 2012; Zwart et al., 2018), and hydrophobic residues were more interchangeable with each other than were polar ones.

The interest in characterizing genotype–fitness maps for other systems was immediately recognized, and experiments have now been performed on genes encoding a diverse range of functions including ubiquitin, poly(A)-binding protein, antibiotic resistance enzymes, a small nucleolar RNA, tRNAs, metabolic enzymes and regulatory regions (Bernet & Elena, 2015; Chan, Venev, Zeldovich, & Matthews, 2017; Dandage et al., 2018; Domingo, Diss, & Lehner, 2018; Firnberg, Labonte, Gray, & Ostermeier, 2014; Jacquier et al., 2013; Klesmith, Bacik, Michalczyk, & Whitehead, 2015; Li, Qian, Maclean, & Zhang, 2016; Li & Zhang, 2018; Melamed, Young, Gamble, Miller, & Fields, 2013; Melnikov, Rogov, Wang, Gnirke, & Mikkelsen, 2014; Puchta et al., 2016; Roscoe, Thayer, Zeldovich, Fushman, & Bolon, 2013; Wrenbeck, Azouz, & Whitehead, 2017). Due to the ease of massively parallel genotype–fitness mapping with this approach, many artificial systems have been devised in which fitness is used as a readout of some other phenotype of interest, such as gene expression (Shultzaberger et al., 2012; Shultzaberger, Malashock, Kirsch, & Eisen, 2010), protein binding affinity (Diss & Lehner, 2018) or protein stability (Kim, Miller, Young, & Fields, 2013). Caution should therefore be taken in comparing studies, as nonlinearities between different phenotypic levels likely have a major influence on the observed properties of the genotype–phenotype relationship, as will be discussed below.

This last decade has thus provided a rich pool of experimental data with which to explore statistically the genotype–phenotype relationship across different scales, and the following will summarize the emerging trends.

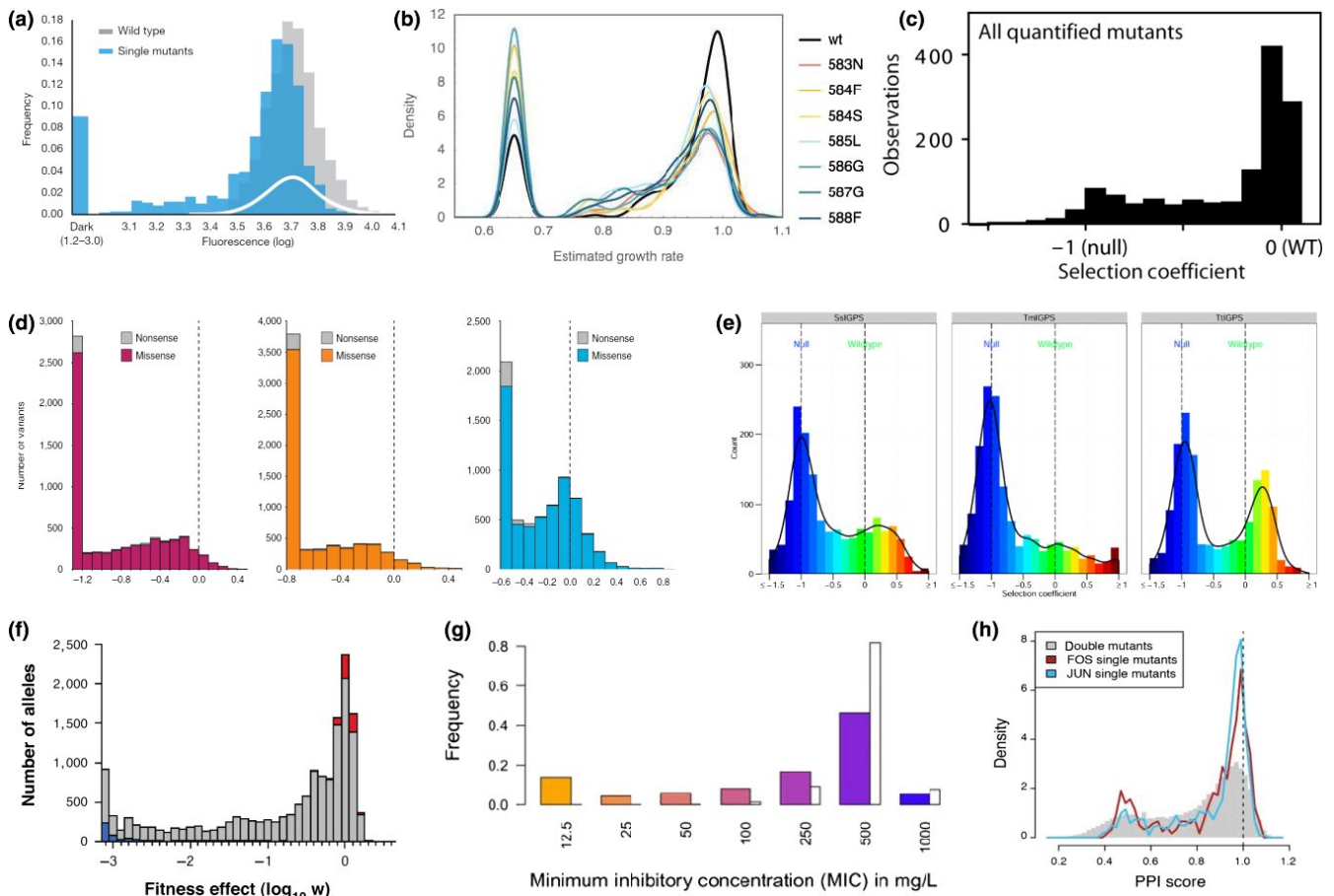## 2.2 | Distribution of mutational effects (DMEs) in single genes

The distribution of effects of individual new mutations is of profound importance to medical and evolutionary genetics (Eyre-Walker & Keightley, 2007). Understanding it can also help guide strategies for directed evolution of useful biomolecules. The deep mutational scanning studies outlined above have revealed, perhaps unsurprisingly, that it differs between coding and noncoding regions, different types of genes and even different regions within genes.

2.2.1 | The DME in single proteins

In proteins, whether the focal phenotype is biochemical functionality (Lagator, Sarikas, Acar, Bollback, & Guet, 2017; McLaughlin et al., 2012; Sarkisyan et al., 2016) or a highly integrated trait like fitness (Bank, Hietpas, Jensen, & Bolon, 2015; Bank, Hietpas, Wong, Bolon, & Jensen, 2014; Chan et al., 2017; Diss & Lehner, 2018; Firnberg et al., 2014; Hietpas et al., 2011; Jacquier et al., 2013; Jiang et al., 2016; Jiang, Mishra, Hietpas, Zeldovich, & Bolon, 2013; Klesmith et al., 2015; Mavor et al., 2016; Melamed et al., 2013; Melnikov et al., 2014; Roscoe et al., 2013; Wrenbeck et al., 2017), the DME appears to be universally multimodal, typically with near-neutral and highly deleterious modes and a vanishing fraction of beneficial effects (Figure 2). An almost trivial exception to this is the DME of repressor proteins on expression (Lagator, Sarikas, et al., 2017), or more generally when a decrease in one phenotype leads to an increase in a downstream phenotype. The same multimodality is found for random mutation samples across whole viral genomes, which are dense with protein-coding sequences (Carrasco, de la Iglesia, & Elena, 2007; Domingo-Calap, Cuevas, & Sanjuán, 2009; Peris, Davis, Cuevas, Nebot, & Sanjuan, 2010; Sanjuan, Moya, & Elena, 2004b). A popular and conceptually simple mechanistic hypothesis for this is as follows: (a) a globally determined property of proteins is stability; (b) stability is potentially affected by amino-acid changes at many positions, while positions contributing to a limiting step in direct protein function are likely to be rare; (c) protein folding is cooperative, resulting in a "thermodynamic cliff" in the stability–folding function; (d) protein mutants therefore tend to lie either at the plateau at the top of this cliff (where the majority of molecules are correctly folded), which is likely where the wild-type resides, or at the bottom of it (majority of molecules unfolded); and (e) the concentration of natively folded molecules is likely a key determinant of protein activity, and the phenotype being measured depends either directly or indirectly on this activity (Bank et al., 2015; Firnberg et al., 2014; Jacquier et al., 2013; Olson, Wu, & Sun, 2014; Sarkisyan et al., 2016; Tokuriki & Tawfik, 2009; Wylie & Shakhnovich, 2011) (Figure 3).

Whatever the mechanism(s) responsible for changes in protein activity, however, the precise form of the distribution of mutational effects must also depend on the quantitative relationship between activity and the measured phenotype, and on the activity of the wild type. For example, three orthologous wild-type indole-3-glycerol phosphate synthases were found to have their neutral mode shifted towards beneficial effects, showing they were sub-optimal for fitness under the chosen experimental conditions (Chan et al., 2017) (Figure 2e). A striking demonstration of the impact of a nonlinear activity–fitness relationship was provided by the Hsp90 chaperone protein (Jiang et al., 2013), found to have a concave, saturating activity–fitness function. Interestingly, similarly shaped functions are found to generally describe activity–*flux* or activity–fitness relationships across enzymes and organisms, in line with expectations from metabolic control analysis (Bershtein, Mu, Serohijos, Zhou, & Shakhnovich, 2013; Dykhuizen, Dean, & Hartl, 1987; Jiang et al., 2013; Kacser & Burns, 1981; Lunzer, Miller, Felsheim, & Dean, 2005) (Figure 4). As is often found, wild-type Hsp90 activity lays safely on this plateau, far from the fitness shoulder, at endogenous expression levels. This meant that even mutations with an intermediate effect on *activity* could be nearly neutral to fitness. By then measuring the distribution of fitness effects at increasingly reduced expression levels, these latent activity effects were increasingly revealed. Indeed, the large number of mutations found to have intermediate, rather than nearly neutral, effects on activity suggested that the dominant mechanism for activity changes was *via* specific molecular function (substrate binding), rather than global stability.

The DME in proteins, as will be found in other sequences, is thus shaped by forces at several scales, from the intramolecular level to the intermolecular interface level to the cell-system level and beyond. A source of bias to be wary of is therefore the choice of experimental system: it is often desirable to focus on systems suspected to show linear, or at least monotonic, relationships across these levels (*e.g.* the activity–fitness function), but more complex relationships may well be common (Baeza-Centurion, Miñana, Schmiedel, Valcárcel, & Lehner, 2019; Hsin-Hung Chou, Delaney, Draghi, & Marx, 2014; Dekel & Alon, 2005; Drummond & Wilke, 2008; Perfeito, Ghozzi, Berg, Schnetz, & Lässig, 2011; Rokyta et al., 2011; Serohijos, Rimas, & Shakhnovich, 2012; Serohijos & Shakhnovich, 2014; Shultzaberger et al., 2010; Towbin et al., 2017). A clear demonstration of this was provided recently for the case of *expression*–fitness relationships— these were characterized in parallel for 81 diverse genes in the yeast, *S. cerevisiae*, by inserting in front of each gene a library of synthetic promoters of known strength (Keren et al., 2016). In addition to the commonly found concave function, a variety of other forms were uncovered, including step-like ones, flat ones and single- or multipeaked ones (Figure 5). Importantly, genes from the same pathway or complex tended to display similar expression–fitness curves, suggesting that these are shaped primarily by the cell-level function of a gene, rather than its specific biochemistry. The consequences of these different elasticity functions for the DME across different genes should therefore be a fruitful avenue for future research.

FIGURE 2 A sample of experimentally characterized distributions of mutational effects (DMEs) in various proteins, all showing at least 2 modes. (a) Distribution of fluorescence intensities resulting from single mutations across the length of a green fluorescent protein (blue). [Reprinted by permission from Springer Nature: *Nature,* Local fitness landscape of the green fluorescent protein (Sarkisyan et al., 2016), Copyright 2016]. (b) Distributions of yeast growth rate effects of all single point mutations in a 9-amino-acid region of 8 variants of a native chaperone protein (the wild-type (black) and 7 single-mutant variants). [Reprinted by permission of the Society for Molecular Biology and Evolution: *Molecular Biology and Evolution, 32*(1), 232, A systematic survey of an intragenic epistatic landscape (Bank et al., 2015)]. (c) Distribution of yeast fitness effects of all single point mutations across the length of native ubiquitin. [Reprinted from *Journal of Molecular Biology, 425*(8), 1,366, Analyses of the effects of all ubiquitin point mutants on yeast growth rate (Roscoe et al., 2013). Copyright (2013), with permission from Elsevier]. (d) Distributions of bacterial fitness effects of all single point mutations across the length of a non-native metabolic enzyme, on which the host strain has been made dependent for nitrogen supply, in the presence of 3 different amide substrates. [Reprinted from (Wrenbeck et al., 2017), licensed under CC BY 4.0]. (e) Distributions of yeast fitness effects of single mutations in the β-barrel core region of 3 phylogenetically divergent orthologues of a metabolic enzyme, on which the host strain has been made dependent for tryptophan biosynthesis. [Reprinted from (Chan et al., 2017), licensed under CC BY 4.0). (f) Distribution of "gene fitness" effects of all single codon substitutions across the length of a native bacterial antibiotic resistance gene, whose product's cellular activity is linked to fitness *via* a synthetic genetic circuit (grey: missense, blue: nonsense, red: synonymous). [Reprinted by permission of the Society for Molecular Biology and Evolution: *Molecular Biology and Evolution, 31*(6), 1583, A comprehensive, high-resolution map of a gene's fitness landscape (Firnberg et al., 2014)]. (g) Distribution of minimum inhibitory concentrations of antibiotic observed for random single point mutations across the length of a native bacterial antibiotic resistance gene (same as f) (coloured bars; white bars: wild type). [Reprinted from *Proceedings of the National Academy of Sciences, 110*(32), 13,068, Capturing the mutational landscape of the beta-lactamase TEM-1 (Jacquier et al., 2013)]. (h) Distributions of complementation assay protein–protein interaction scores resulting from single point mutations in the leucine zipper domains of 2 human transcription factor subunits (red and blue). [Reprinted from (Diss & Lehner, 2018), licensed under CC BY 4.0]

2.2.2 | The DME in single functional noncoding RNAs

The DME for the few functional noncoding RNA sequences (*cis*-regulatory region, tRNA, snoRNA, twister ribozyme) so far examined appears to follow similar rules to proteins: deleterious and nearly neutral modes and a miniscule proportion of beneficial mutations

again probably shaped in part by both a thermodynamic stability threshold and common saturating concave activity–fitness functions (Bendixsen, Østman, & Hayden, 2017; Bernet & Elena, 2015; Kobori & Yokobayashi, 2016; Li et al., 2016; Puchta et al., 2016). This is perhaps not surprising because (a) as for proteins, noncoding RNA function depends strongly on structure, and (b) there is no obvious reason why the dependence of cell fitness on RNA-regulated
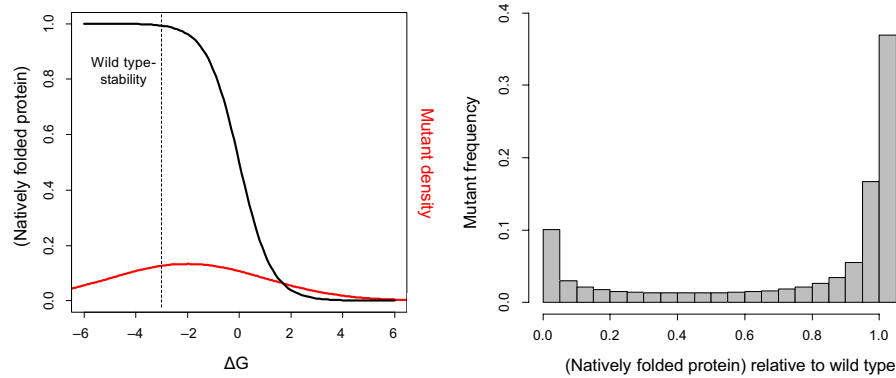
FIGURE 3 Illustration of the thermodynamic hypothesis for DMEs in proteins. Left panel—Black sigmoid curve shows the fraction of natively folded protein molecules as a function of the free energy of folding, ΔG, following: $P^{nat} = \frac{1}{1+e^{\Delta G/k_bT}}$, where k_b is the Boltzmann constant and T is temperature (k_bT is set here to 0.62, as in (Wylie & Shakhnovich, 2011)). Dashed line marks a hypothetical wild-type protein stability (−3 kcal/mole), located on the plateau of the sigmoid, for illustration. Red curve shows a hypothetical distribution of mutant ΔG values, resulting from a DME on ΔG that is Gaussian with a mean of +1, following (Wylie & Shakhnovich, 2011), but here with a larger standard deviation of 3. The stability sigmoid could be steepened by effects such as irreversible aggregation or degradation of misfolded species (Tokuriki & Tawfik, 2009). Right panel: The resulting DME on the *relative fraction of natively folded protein molecules*, which is bimodal under these parameter values

activities or direct RNA activities (especially those chosen for ex-perimental investigation) would be fundamentally different to its dependence on protein activities.

2.2.3 | The DME in single cis-regulatory DNA regions

In the case of *cis*-regulatory DNA sequences, measured DMEs are generally unimodal (Badis et al., 2009; Kinney et al., 2010; Lagator, Igler, Moreno, Guet, & Bollback, 2016; Lagator, Sarikas, et al., 2017; Warren et al., 2006), although they can be multimodal in more complex regulatory contexts (Lagator, Paixão, Barton, Bollback, & Guet, 2017; Lagator, Sarikas, et al., 2017) or when a fraction of the sites function through specific base-pairing with other nucleic acids (and so are highly sensitive to mutation; Boyle et al., 2017). The majority of mutations decrease DNA binding

to regulators (again presumably because wild-type sequences are optimized for relatively strong binding), but of course the im-pact of this on downstream phenotypes depends if the regula-tion is activational, repressive or some complex mixture of both (Lagator, Paixão, et al., 2017). The biophysical reason for a more uniform DME on direct biochemical phenotypes for *cis*-regula-tory DNA sequences than for proteins and noncoding RNAs is not clear: the thermodynamics of DNA–protein binding can result in similar energy–phenotype functions to those of macromolecular folding described earlier (Lagator, Paixão, et al., 2017; Mustonen, Kinney, Callan, & Lassig, 2008; Vilar, 2010). It may be that in re-ality they are less steep (e.g. because deleterious effects from misfolded species no longer contribute) or that *cis*-regulatory mu-tations tend to have smaller effects on binding energy than do protein and noncoding RNA mutations on folding energy, and so sample a smaller region of the energy space (see Figure 3).

FIGURE 4 A sample of experimentally characterized elasticity functions, all of a saturating concave form. (a) Function: expression level—growth rate; protein: chaperone; organism: *Saccharomyces cerevisiae*. [Reprinted from (Jiang et al., 2013), licensed under CC BY 4.0]. (b) Function: enzymatic performance (k_{cat}/K_m)—fitness, under 2 different coenzymes; protein: oxidoreductase (amino-acid biosynthesis); organism: *Escherichia coli*. [From *Science*, *310*(5,747), 501, The biochemical architecture of an ancient adaptive landscape (Lunzer et al.., 2005). Reprinted with permission from AAAS]. (c) Left-right, top-bottom. Functions: enzymatic activity—metabolic flux (first 4), gene dose—growth rate, enzymatic efficiency (V_{max}/K_m)—metabolic flux, gene dose—DNA repair rate, enzymatic efficiency—metabolic flux; proteins: lyase, transferase, ligase, aminotransferase (all from same amino-acid biosynthesis pathway), carboxylase (nucleotide biosynthesis), oxidoreductase (melanin biosynthesis), unknown gene defective in a *xeroderma pigmentosum* patient (nucleotide repair), oxidoreductase (ethanol oxidation); organism: *Neurospora crassa* (first 4), *Saccharomyces cerevisiae*, *Mus musculus*, *Homo sapiens*, *Drosophila melanogaster*. [Republished with permission of Genetics Society of America, from *Genetics*, *97*(3–4), 642, The molecular basis of dominance (Kacser & Burns, 1981); permission conveyed through Copyright Clearance Center, Inc.]. (d) Function: expression level—growth rate; protein: oxidoreductase (cofactor biosynthesis); organism: *Escherichia coli*. [Reprinted from *Molecular Cell*, *49*(1), 137, Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness (Bershtein et al., 2013). Copyright (2013), with permission from Elsevier]. (e) Function: enzymatic activity—fitness; proteins: sugar:proton symporter and hydrolase (sugar catabolism); organism: *Escherichia coli*. [Republished with permission of Genetics Society of America, from *Genetics*, *115*(1), 29, Metabolic flux and fitness (Dykhuizen et al., 1987); permission conveyed through Copyright Clearance Center, Inc.]

$$y = x/(x + 0.014)$$
$$R^2 = 0.98$$
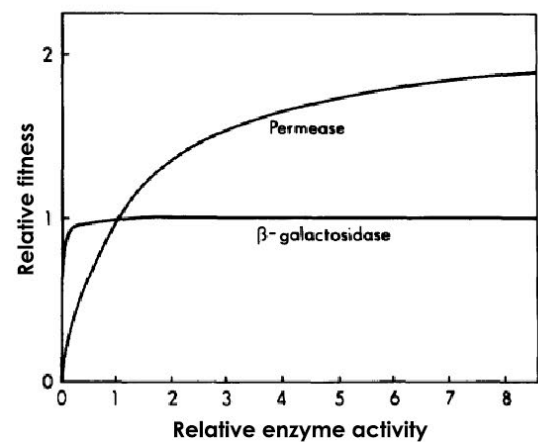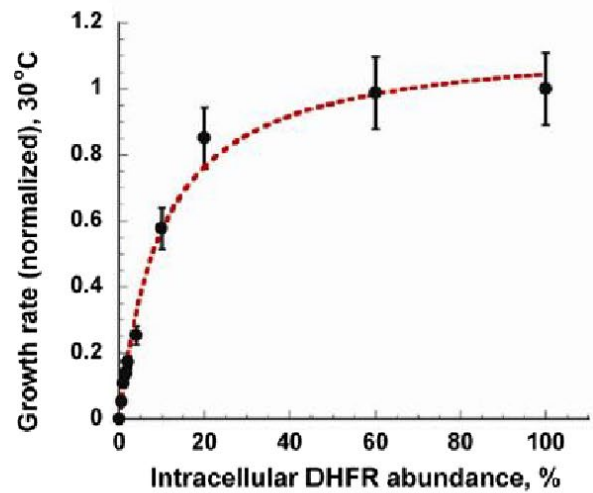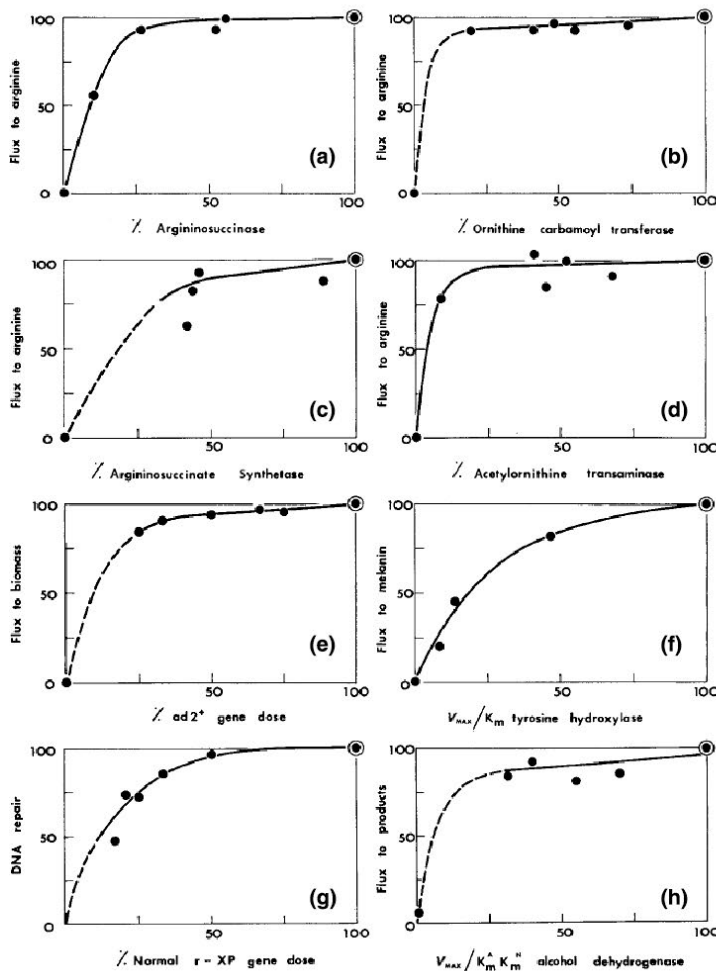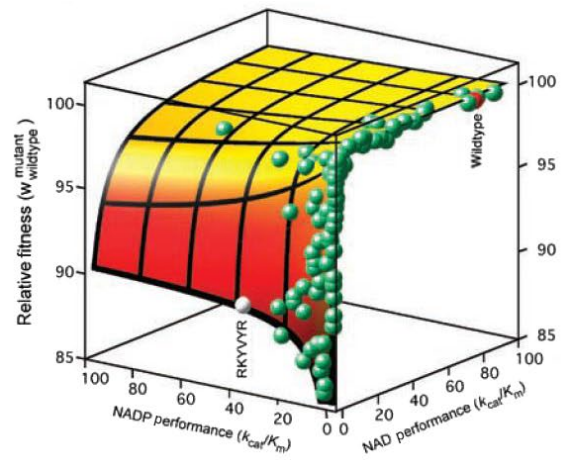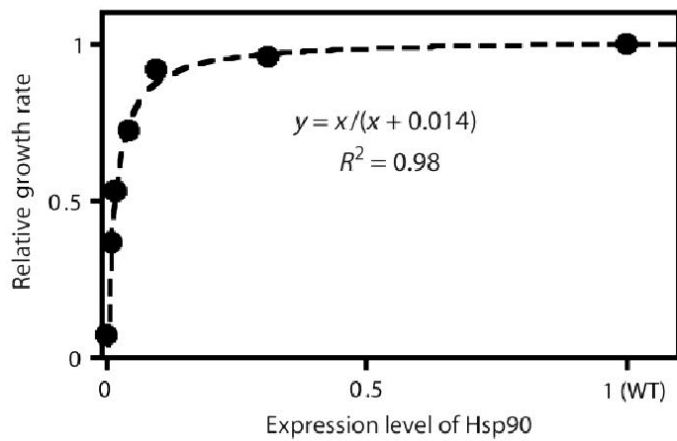
2.3 | Epistasis within single genes (intragenic epistasis)

Epistasis is defined here as the deviation of an observed phenotype value from that expected if the constituent individual mutations combined additively on the log scale (*i.e.* multiplicatively on the linear scale; Wolf, Brodie III, & Wade, 2000) (Figure 6). Epistasis is critically important in medical and evolutionary genetics and bioengineering: among

other things, it confounds prediction of mutational effects, constrains adaptive paths, determines the benefit of sex and hinders efforts to increase yields and activities of industrially useful substances (Badano & Katsanis, 2002; Breen, Kemena, Vlasov, Notredame, & Kondrashov, 2012; Dipple & McCabe, 2000; Hansen, 2013; Kimura & Maruyama, 1966; Kondrashov, 1988; Kondrashov & Kondrashov, 2001; Manolio et al., 2009; Niederberger, Prasad, Miozzari, & Kacser, 1992; Phillips, 2008; Scriver & Waters, 1999; Weinreich, 2006). Intragenic epistasis

is logically shaped by the same forces as those shaping the DME, such as thermodynamics and phenotype–phenotype functions (Lehner, 2011; de Visser, Cooper, & Elena, 2011), and its basic statistical properties will now be summarized for the systems in which large-scale measurements have been made.

### 2.3.1 | Intragenic epistasis within single proteins

Within proteins, epistasis appears prevalent and predominantly negative, with a unimodal distribution. As the majority of mutations tend to reduce the value of the measured phenotype, this epistasis mainly represents *synergistic* interactions between deleterious mutations: the combined impact of multiple mutations is often *worse* than "the sum of the parts" (Bank et al., 2015; Bank, Matuszewski, Hietpas, & Jensen, 2016; Melamed et al., 2013; Olson et al., 2014; Sarkisyan et al., 2016) (but Araya et al., 2012; Diss & Lehner, 2018 are exceptions). Under the stability threshold model outlined above, such synergistic negative epistasis would indeed be expected to prevail between mutations having mildly destabilizing effects, because, beginning from the stability plateau, the downward slope initially becomes steeper as stability is decreased. After a threshold is crossed, however, the slope levels off again, which can result in positive epistasis being detected between highly destabilizing mutations if the protein is nonessential or if an experimental measurement limit is approached (Bank et al., 2015; Bershtein,

Segal, Bekerman, Tokuriki, & Tawfik, 2006; Diss & Lehner, 2018; Jacquier et al., 2013; Sarkisyan et al., 2016; Wylie & Shakhnovich, 2011; Figure 7). Differences in the destabilizing effects of mutations, wild-type stability and measurement precision/range could therefore explain discrepancies between studies as to the pervasiveness of epistasis as well as the relative fractions of positive and negative interactions. Negative epistasis between mutations of mildly deleterious biochemical effect could also result from concave saturating elasticity functions (described above) (Bank et al., 2015; Szathmary, 1993), for the same reason that it results from the stability function, and analogously to the classical molecular hypothesis of genetic dominance and recessivity (Kacser & Burns, 1981). Of course, as for the DME, many specific cases will deviate from these general expectations and vary across proteins. For example, specific structural interactions could generate sign epistasis when mutations do not act additively on the level of *folding energy* (ΔG) (Otwinowski, McCandlish, & Plotkin, 2018; Starr & Thornton, 2016). Further, any single mutation could potentially effect multiple molecular phenotypes, including protein stability, affinity, activity (existing or new) per folded species, folding and assembly kinetics, aggregation propensity, degradation rate and post-translational modification, as well as RNA splicing (Baeza-Centurion et al., 2019; DePristo, Weinreich, & Hartl, 2005; Echave & Wilke, 2017; Otwinowski, 2018; Shah, McCandlish, & Plotkin, 2015; Sikosek & Chan, 2014), so simplistic models should be treated with caution.
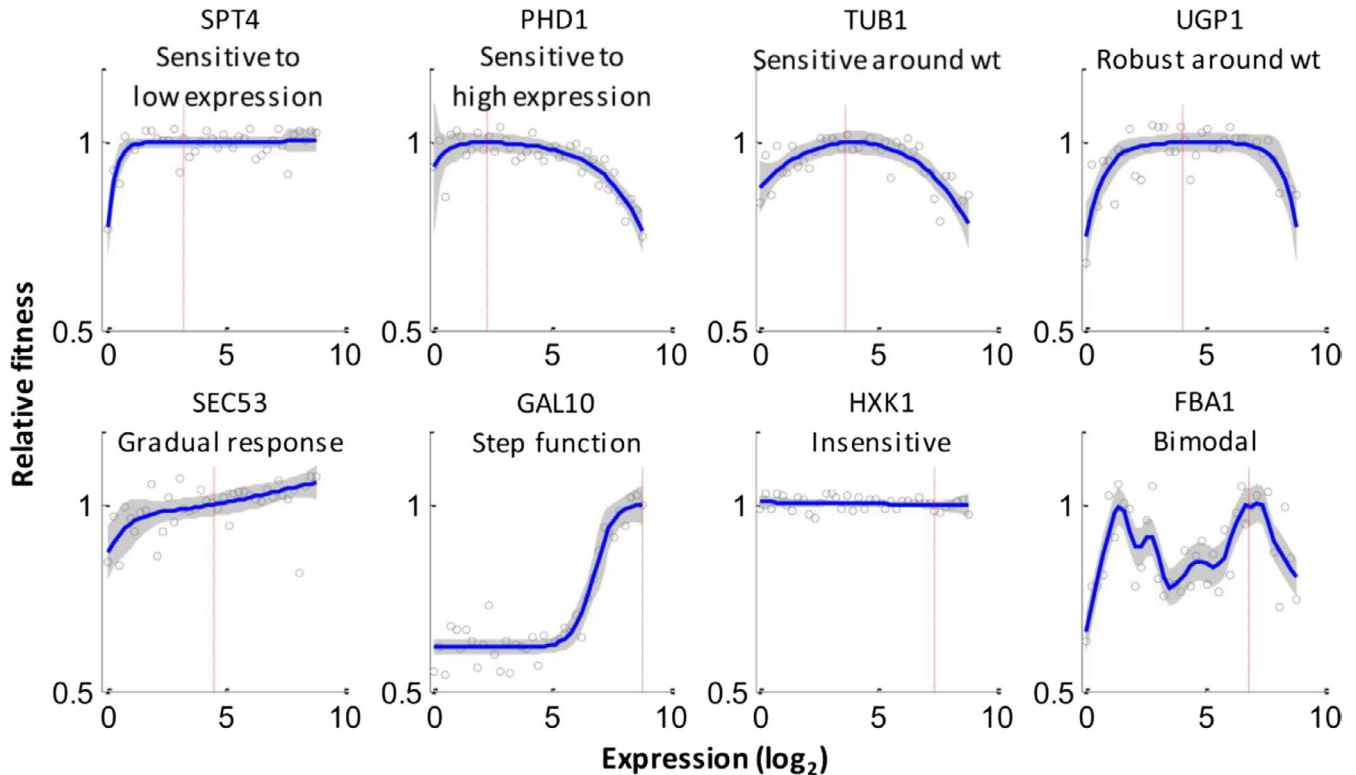
FIGURE 5 Expression–fitness functions for a diverse set of protein-coding yeast genes. Red lines mark wild-type expression levels. [Reprinted from *Cell*, *166*(5), 1,286, Massively parallel interrogation of the effects of gene expression levels on fitness (Keren et al., 2016). Copyright (2016), with permission from Elsevier]
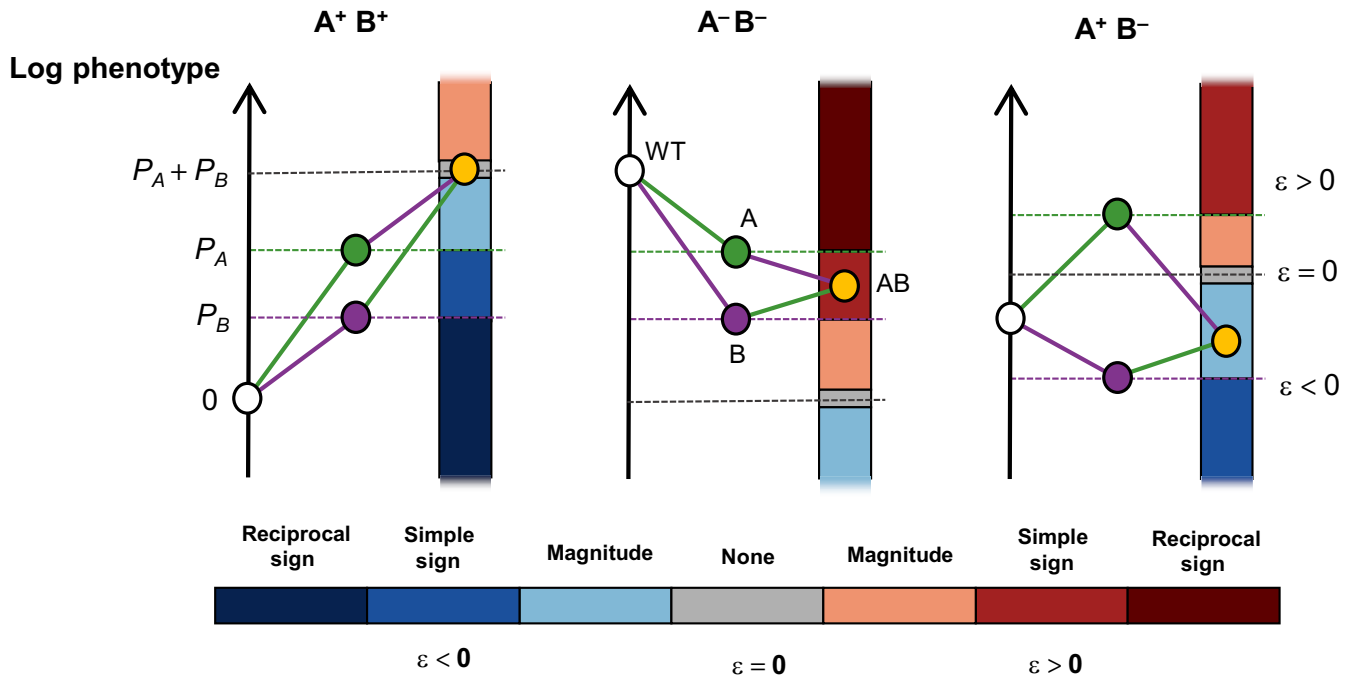
FIGURE 6 Categories of epistasis possible for different types of mutation pairs. "A" and "B" are mutations, and superscript "+" and "−" denote that these individual mutations increase or decrease the value of the measured phenotype, P. In all cases, the white point is wild type and the orange point is the AB double mutant. Grey dashed line marks the sum of P_A and P_B, that is the *expected* value for the double mutant. Epistasis measures the deviation from this expectation, which may be either negative or positive, and can be categorized as either magnitude (the direction of mutational effects do not depend on the other mutation) or sign type. Sign epistasis can be further categorized as simple (effect of one mutation changes sign in the presence of the other) or reciprocal (effects of both mutations change sign in the presence of the other). The three examples shown are (left-right): no epistasis between a pair of positive-effect mutations, positive simple sign epistasis between a pair of negative-effect mutations, and negative magnitude epistasis between a positive-effect and negative-effect mutation

Indeed, a recent study demonstrated how mRNA secondary structure phenotypes affecting translation efficiency can have fitness consequences through a general, cell-level mechanism of ribosome sequestration (Cambray, Guimaraes, & Arkin, 2018).

In addition to prevailing negative epistasis, a trend of "increasing losses" is sometimes detected in proteins (at least for mildly deleterious mutations), further supporting the idea that there exists some source of concavity in the phenotypic landscape (Bank et al., 2015; Diss & Lehner, 2018; Sarkisyan et al., 2016). This trend manifests as a positive correlation between the magnitude of individual mutation effects and the magnitude of epistasis they experience when combined. The flip side of "increasing losses" between the predominantly deleterious mutations of the DME is "diminishing returns" between the rare mutations from the beneficial tail of the DME (MacLean, Perron, & Gardner, 2010; Nagel, Joyce, Wichman, & Miller, 2012; Schenk, Szendro, Salverda, Krug, & de Visser, 2013): synergistic negative epistasis increases with the *increasing* downward slope moving away from the plateau, and antagonistic negative epistasis increases with the *decreasing* upward slope. The evolutionary outcome of these trends should be a more rapid purging of deleterious mutation combinations (negative selection), but a slower rate of adaptation resulting from beneficial combinations (positive selection), than would be expected in the absence of epistasis.

## 2.3.2 | Intragenic epistasis within functional noncoding RNAs

Epistasis between mutations residing in the same noncoding RNA molecule, as for the DME and again probably for the same reasons, appears similar to the case of proteins: it is common and usually biased towards negativity (Bendixsen et al., 2017; Domingo et al., 2018; Kobori & Yokobayashi, 2016; Li et al., 2016; Chuan Li & Zhang, 2018; Puchta et al., 2016). These experimental results are in contrast to the predominance of positive epistasis predicted by earlier computational RNA folding models (Wilke & Christoph, 2001; Wilke, Lenski, & Adami, 2003), but it is not clear whether this is due to the binary nature of these models or the fact that they do not use naturally evolved sequences as their starting point (Bendixsen et al., 2017). When higher-order epistasis was examined experimentally, positive and negative epistasis were found to be equally prevalent, with many cases of sign epistasis, suggesting a rugged multidimensional fitness landscape (Domingo et al., 2018; *cf.* Bank et al., 2016) for a protein region).

2.3.3 | Epistasis within single cis-regulatory DNA regions

Epistasis within *cis*-regulatory DNA has also been studied experimentally in a limited number of contexts. Within a mammalian

Rhodopsin promoter region bound by at least two transcription factors, epistasis for expression was found to be biased towards negativity (Kwasnieski, Mogno, Myers, Corbo, & Cohen, 2012). More strikingly, in a recent study of the direct effect of target-site mutations on dCas9-DNA binding, negative epistasis for initial association rate between the universally deleterious single mutations was found to be ubiquitous—binding was essentially always worse than would be predicted from the simple addition of individual mutation effects (Boyle et al., 2017). Cas9-DNA binding may be rather unrepresentative, however, as its function is clearly not regulatory (it is immune), and it is mediated by nucleic acid base-pairing. Another team recently examined the expression epistasis between mutations within both the same and different operators in two different promoters: the *E. coli araBAD* promoter (Lagator et al., 2016) and the lambda bacteriophage P_R promoter (Lagator, Paixão, et al., 2017). Although these two promoters possess substantially different architectures, both exhibited a predominance of negative epistasis, whether mutations resided in the same or different operators and whether active repressor proteins were present in high concentration or not. For the simple case where expression is determined by DNA binding to a single regulator, this is to be expected from thermodynamic considerations: the free binding energy–expression curve resulting from their generic model is, once again, of a concave, saturating form (Lagator, Paixão, et al., 2017). Notably, though, in the lambda bacteriophage P_R promoter, the fraction of positive interactions increased when active repressor concentration was increased, as did the frequency of the most extreme form of interaction, reciprocal sign epistasis (from 8% to 66%). This is explained by the fact that repressor binding sites (operators) and RNA polymerase binding sites overlap in this promoter, and so, promoter mutations will tend to effect binding to both of these—one of which decreases transcription, the other of which increases transcription (Lagator, Paixão, et al., 2017).

Cis-regulatory DNA sequences have thus proved to be an excellent model system for studying how epistasis can arise through the inherent *molecular* pleiotropy of mutations: they directly affect multiple molecular phenotypes (here, binding energy with different regulatory proteins), often differentially, and the *measured* phenotype (here, expression) is then some function of these multiple input phenotypes. In reality, as for proteins and noncoding RNAs, many more molecular phenotypes than those typically considered are potentially impacted by an individual mutation, such as DNA structure (Rajkumar, Dénervaud, & Maerkl, 2013), binding site accessibility (Levo & Segal, 2014), regulator protein cooperativity (Todeschini, Georges, & Veitia, 2014) and long-range DNA looping (Levine, Cattoglio, & Tjian, 2014), likely accounting for the significant fraction of observed interactions that are not explained by simple binding energy models (Lagator, Paixão, et al., 2017). Current empirical studies thus present a minimal mechanistic picture of the potential sources of epistatic interactions, and the goal now should be to find additional molecular phenotypes that can account for some of the unexplained epistasis.
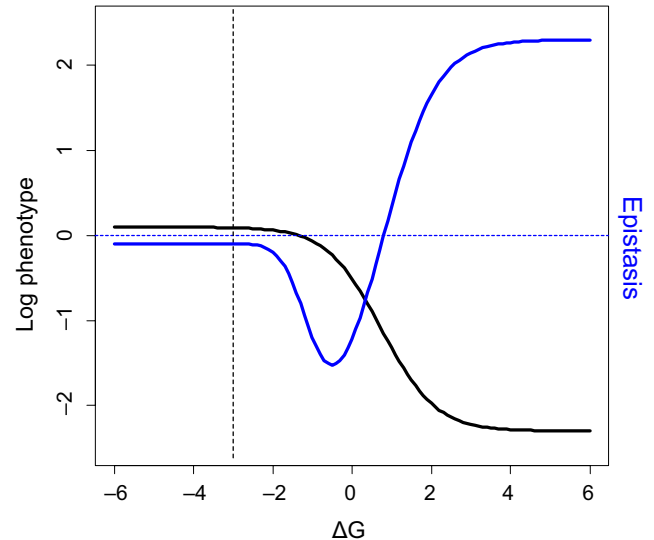
FIGURE 7 Trends of epistasis predicted by thermodynamic model of mutation effects. Black sigmoid curve shows the logarithm of a phenotype that increases proportionally with the fraction of natively folded protein molecules as a function of the free energy of folding, *ΔG*. A small background value (phenotype of 0.1 in the absence of any correctly folded molecules) has been applied to capture the situation for nonessential genes and/or the effect of measurement background/limits. If the phenotype truly approaches 0 in the limit of very high *ΔG*, the stability curve is no longer sigmoidal on this log scale, but has a concave shape, causing epistasis (see below) to become increasingly negative, in a linear fashion, as *ΔG* increases. In reality, however, biology or experimental limitations often result either in a background phenotype value in the absence of correctly folded protein, resulting in a log-sigmoid as shown here, or in a threshold being applied below which all mutants are considered null and therefore not considered for epistasis analysis. The formula and parameter values are as for Figure 3, with the addition of the 0.1 phenotype background. Dashed vertical line marks a hypothetical wild-type protein stability (−3 kcal/mole), located on the stability plateau. Blue curve shows the epistasis that would occur between pairs of mutations of identical *ΔG* effects, each of which individually displaces *ΔG* from the wild-type value to the value indicated by the x-axis. Dashed horizontal line marks the boundary between positive and negative epistasis (*i.e.* zero epistasis). A transition from negative to positive epistasis occurs as mutations become more strongly destabilizing, due to the sigmoidality of the stability curve. The shape of the epistasis curve could explain why both negative and positive epistasis are observed between mutations within proteins, as well as the existence of certain correlations between mutation effect size and epistasis (see below)

Finally, *cis*-regulatory sequences appear to play a major role in evolutionary processes (Wittkopp & Kalay, 2012; Wray, 2007), yet none of the studies above have assessed epistasis at the level of fitness. As for proteins and RNA, in addition to the mechanisms just discussed, this is bound to be also shaped by the expression–fitness elasticity functions of the regulated genes. Fortunately, these are far more accessible to high-throughput genome-wide characterization than are pure activity–fitness functions (Keren et al., 2016).

3 | TWO GENES

Adding just one more gene, let alone several, to the kinds of high-throughput studies that have provided such rich insights for single genes generally requires more than simply twice the effort ("experimental effort epistasis"). For many systems, the problem is that the length of DNA fragments required to carry two different genes exceeds those amenable to the highest-throughput sequencing technologies (max. ~1kb with Illumina). Presumably largely for this reason, very few deep mutational scanning studies have assayed more than one gene or even regulatory sequence simultaneously, but unique DNA barcodes that allow a long sequence to be broken up for sequencing and then reassembled in silico provide one workaround (Fowler, Stephany, & Fields, 2014; Sarkisyan et al., 2016).

3.1 | Epistasis between two genes (intergenic epistasis)

3.1.1 | Intergenic epistasis between two physically interacting partners

A recent exception was the measurement of intermolecular epistasis between a library of point mutants of the leucine zipper domains of two proto-oncogenes, FOS and JUN (Diss & Lehner, 2018). The products of these two genes physically associate through these domains to form a transcription factor complex, AP-1. In this case, the technical challenge of long sequencing fragments was overcome by isolating the two small leucine zipper domains, which are presumed to function in a modular manner, and cloning them adjacent to each other in the absence of the rest of the native proteins. An artificial complementation assay, in which intermolecular FOS-JUN binding drives the assembly of a drug-resistance enzyme, was used to link binding strength to cell growth in the presence of the drug. The relationship between abundance of the complementation complex and growth rate is well-characterized and expected to be approximately linear. Intermolecular epistasis was found to be common and just slightly biased towards negativity. Importantly though, a characteristic relationship between the effect of single mutations and the interaction they experienced when combined suggested that the epistasis could be partly explained by a sigmoidal thermodynamic fitness landscape describing intermolecular binding (Figure 7). Indeed, applying this model removed the correlation between individual effects and epistasis, and increased the percentage of explained variance in double-mutant phenotype scores from ~86% (under a simple multiplicative model) to ~89%. It may, however, be a particularity of the system, as the folding of leucine zippers like FOS and JUN is known to be coupled to their binding (Patel, Abate, & Curran, 1990; Thompson, Vinson, & Freire, 1993); cf. the 3-state biophysical model of (Otwinowski, 2018)—studies on other pairs of binding partners are therefore necessary.

An impressive effort was also recently made to characterize the epistasis between a large number of mutants of a particular repressor protein and its cis-regulatory DNA target at low-throughput, using a fluorescent reporter to measure expression (Lagator, Sarikas, et al., 2017). Epistasis was detected for about half of the 150 pairwise interactions tested, the majority of which were positive. As before, much, but not all, of this epistasis could be rationalized in terms of the promoter architecture, which contained overlapping binding sites for the mutated repressor protein and RNA polymerase, raising hope that a set of rules may exist that adequately predicts epistasis both within promoters, and between promoters and regulators, when promoter architecture is known.

3.1.2 | Intergenic epistasis between two functionally interacting partners

Although epistasis between protein–protein and protein–DNA binding partners is of great importance, the majority of genes in a given genome are expected to interact indirectly, through metabolic, regulatory and signalling networks. Metabolic networks are the most tractable of these, being based on simple mass flow. Metabolic control analysis (MCA) provides a rigorous framework to explore how pathway phenotypes such as flux and metabolite concentrations depend on the activity of several enzymes simultaneously, and thus enables predictions of interenzyme epistasis (Szathmary, 1993). Several small-scale studies provide support for the validity of MCA in general (Hsin-Hung Chou et al., 2014; Dykhuizen et al., 1987), but its rich predictions regarding the epistasis between genes connected through metabolic pathways remain untested. Importantly, the nature of fitness epistasis is predicted to vary considerably depending on which pathway phenotypes are under selection (flux, steady-state metabolite levels), the pathway position of any selected metabolites relative to the two enzymes considered and the type of selection operating (directional, stabilizing) (Szathmary, 1993) (Figure 8). This points to the necessity of uncovering which phenotypes are typically under selection if we are to use such systems models to predict sequence–fitness relationships. Epistasis between two or more genes in signal-flow networks has also never been tested on a systematic scale, but a recent small-scale study on a synthetic gene regulatory cascade found a surprisingly high frequency of sign epistasis simply at the level of expression (briefly, explained by the fact that changes in the activity of one regulator shift the optimal activity of other regulators) (Nghe, Kogenaru, & Tans, 2018).

4 | THE GENOME

4.1 | Experimental approaches for genome-wide genotype–phenotype mapping

Scaling up deep mutational scanning experiments to the scale of the genome is at present out of reach: bottlenecks include the precise, genome-wide introduction of individual mutations (mutagenesis

efficiency and accuracy), sequencing costs and associating mutations at distant loci. The first is improving with advances in genome engineering, particularly from CRISPR-Cas9-based methods (Barbieri, Muir, Akhuetie-Oni, Yellman, & Isaacs, 2017; Haimovich, Muir, & Isaacs, 2015; Roy et al., 2018). The second is continuing to improve, following a long-term trend of decreasing costs (but see (Muir et al., 2016) for the alternative challenge of managing increasing amounts of data); and the third is becoming more feasible with emulsion-based generalized DNA assembly technologies that encapsulate single cells and enable distal DNA sites to be linked by sequencing (either by directly ligating mutated sites adjacent to each other (Haliburton, Shao, Deutschbauer, Arkin, & Abate, 2017; Zeitoun et al., 2015) or, more scalably, by ligating them to a cell-specific DNA barcode (Zeitoun, Pines, Grau, & Gill, 2017)) (Figure 9).

In the meantime, a plethora of "functional genomics" systematic genome-wide studies have been performed, especially in yeast, that measure the effects (typically growth) of different types of large perturbations in single or multiple (maximum of 3) genes (deletion, overexpression, knockdown, transposon insertion; typically one or two perturbations per gene) (Baba et al., 2006; Babu et al., 2014; Boutros et al., 2004; Breslow et al., 2008; Collins et al., 2007; Costanzo et al., 2010, 2016; Davierwala et al., 2005; Decourty et al., 2008; Douglas et al., 2012; Fuchs et al., 2010; Gagarinova et al., 2016; Giaever et al., 2002; Jaffe, Sherlock, & Levy, 2017; Kamath et al., 2003; Kuzmin et al., 2018; Nichols et al., 2011; Onge et al., 2007; van Opijnen, Bodi, & Camilli, 2009; Roguev et al., 2008; Sameith et al., 2015; Schuldiner et al., 2005; Szappanos et al., 2011; Tischler, Lehner, Chen, & Fraser, 2006; Tong, 2004; Tsherniak et al., 2017; Ursell et al., 2017). These provide rich functional data sets, but there is little evidence that any genotype–phenotype inferences would generalize to the less extreme perturbations (*e.g.* point mutations) often found in nature (Rojas Echenique, Kryazhimskiy, Nguyen Ba, & Desai, 2019). Further, the majority are biased towards altering functions of known genes. An approach to study genome-wide genotype–fitness relationships at the other extreme is mutation accumulation experiments and the analysis of natural DNA sequence data. These benefit from sampling *naturally occurring* mutations, which may be very different to those introduced experimentally, and potentially capturing mutations of effects too weak to be quantified directly, but they rely purely on inferences (Eyre-Walker & Keightley, 2007). In between these two approaches are experiments that directly measure the effects of randomly induced or collected mutations (Bonhoeffer, 2004; Carrasco et al., 2007; Domingo-Calap et al., 2009; Peris et al., 2010; Sanjuan et al., 2004b; Szafraniec, Wloch, Sliwa, Borts, & Korona, 2003; Wloch, Szafraniec, Borts, & Korona, 2001), or of mutations that are detected during experimental evolution (Caudle, Miller, & Rokyta, 2014; Chou, Chiu, Delaney, Segre, & Marx, 2011; Chou et al., 2014; Flynn, Cooper, Moore, & Cooper, 2013; Khan, Dinh, Schneider, Lenski, & Cooper, 2011; Kryazhimskiy, Rice, Jerison, & Desai, 2014; Kvitek & Sherlock, 2011; Rokyta et al., 2011; Venkataram et al., 2016), all of which provide data sets orders of magnitude smaller than do systematic perturbation studies.

4.2 | The genome-wide DME

An early finding of the genome-wide perturbation studies was that, in most organisms and in permissive conditions, the majority of genes are inessential (Baba et al., 2006; Gerdes et al., 2003; Giaever et al., 2002; D.-U. Kim et al., 2010; Sassetti, Boyd, & Rubin, 2003; Viswanatha, Li, Hu, & Perrimon, 2018; Yamamoto et al., 2009; Zhang & Lin, 2009). A notable exception is the "minimal bacterium", *Mycoplasma genitalium* (Glass et al., 2006; Hutchison et al., 1999), which was in fact the first organism in which gene essentiality was examined directly (Hutchison et al., 1999), demonstrating a frequent irony in experimental biology: the first choice of experimental system is usually based on convenience, which in some cases results in it being utterly unrepresentative. "Chemical genomics" approaches, which phenotype genome-wide perturbation libraries in many different defined environments, perhaps unsurprisingly, reduce the fraction of inessential genes by revealing *conditionally essential* genes that are necessary for growth in at least one of the environments tested (Nichols et al., 2011). The interpretation of these classifications (essential, conditionally essential, inessential) is not clear, however, as they are clearly fully dependent on the environments tested (see *Environment*).

Interestingly, it seems that the DFE of single-gene deletions/disruptions (Baryshnikova et al., 2010; van Opijnen, Lazinski, & Camilli, 2014; Wang et al., 2018) might be qualitatively similar to the DFE of randomly sampled/induced genome-wide mutations (Eyre-Walker & Keightley, 2007), itself similar to the DFE of single point mutations in proteins and noncoding RNAs. These all comprise a nearly neutral mode with a heavy negative tail, a very deleterious/lethal mode (sometimes ignored) and a small proportion of beneficial effects (see (Bataillon & Bailey, 2014; Eyre-Walker & Keightley, 2007) for more fine-scale properties). Uncovering the reasons for this universality should prove to be a fascinating endeavour.

In the meantime, top-down heuristic phenotype–fitness models have provided useful unifying frameworks with which to capture such common trends found across these different scales and species, as they do not rely on system-specific mechanistic details. In particular, Fisher's geometric model of adaptation (FGM), originally proposed simply as a convenient metaphor for phenotypic adaptation (Fisher, 1930), can correctly predict the oft-observed shifted reflected Γ-shape of the nearly neutral mode of the DFE (although not generally the strongly deleterious/lethal mode) (Bank et al., 2014; Bataillon & Bailey, 2014; Chevin, Martin, & Lenormand, 2010; Jacquier et al., 2013; Martin, 2014; Martin & Lenormand, 2006; Tenaillon, 2014; Trindade, Sousa, & Gordo, 2012). In addition, FGM predicts the common patterns of epistasis observed between beneficial mutations in systems of all scales (see also below): a general predominance of negative (antagonistic) epistasis and, more specifically, a trend of diminishing returns (Blanquart, Achaz, Bataillon, & Tenaillon, 2014; Martin, Elena, & Lenormand, 2007; Rokyta et al., 2011). These agreements are not so surprising when we consider the
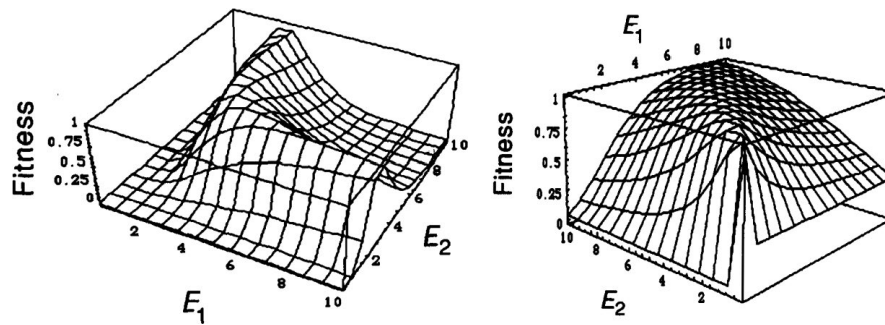
FIGURE 8 Two-enzyme activity–fitness functions predicted from metabolic control analysis. E_1 and E_2 are the activities of two enzymes acting at adjacent steps of a linear metabolic pathway. In both plots, fitness is assumed to depend solely on the steady-state concentration of a pathway intermediate, in a Gaussian manner (*i.e.* stabilizing selection is assumed to operate on the intermediate). The only difference is that, in one case, the intermediate lies downstream of the two enzymes (left), and in the other, it lies between them (right). The two landscapes have strikingly different forms, resulting in different expectations of interenzyme epistasis. Further, in both cases, trends of interenzyme epistasis will depend on the position of the wild type and the distribution of mutation effects on enzyme activities. [Republished with permission of Genetics Society of America, from *Genetics*, *133*(1), 129–130, Do deleterious mutations act synergistically? Metabolic Control Theory provides a partial answer (Szathmary, 1993); permission conveyed through Copyright Clearance Center, Inc.]
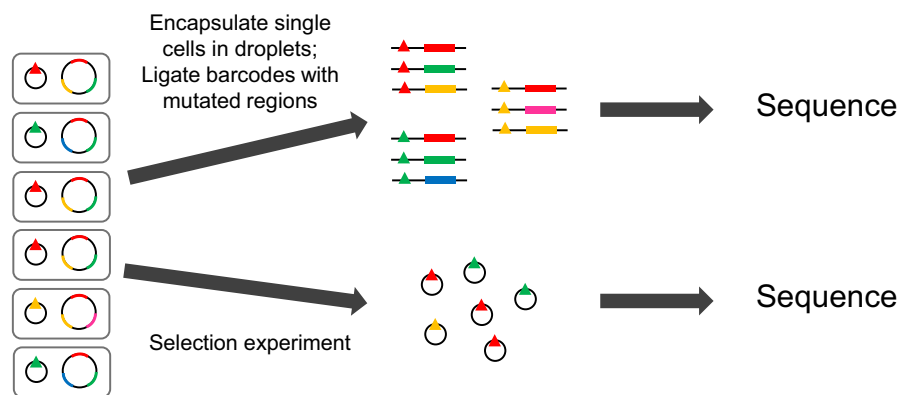
FIGURE 9 Barcoded-Tracking of Combinatorial Engineered Libraries (bTRACE), a general high-throughput method for analysing the effect of known genome-wide mutation combinations. A multiplex genome-engineering method is used to construct a cell library in which each clone can contain multiple mutations throughout the genome, and this library is itself transformed with a library of plasmids carrying highly diverse DNA barcodes (triangles), such that each cell now contains a unique barcode. Single cells are then encapsulated in emulsion droplets, where they are lysed, and a targeted binary PCR assembly reaction is performed to ligate barcodes adjacent to chosen genomic regions. The emulsion is broken, and deep-sequencing of the assembled product pool allows reconstruction of the complete phased genotype associated with each barcode. In parallel, the library can be phenotyped by one of the deep-sequencing techniques discussed previously, with only the small DNA barcodes now requiring sequencing, allowing genome-wide mutation combinations to be linked to an amenable trait at high throughput. Figure based on Figure 1 from Zeitoun et al. (2017)

similarities of a Gaussian FGM to the sigmoidal and simple concave phenotype–fitness functions previously demonstrated to explain these patterns: in all cases, as the fitness maximum is approached, the upward slope becomes increasingly flat (*i.e.* they are all locally concave at high fitness). One difference is that, as opposed to the sigmoidal and simple concave functions, the nonmonotonicity of FGM also predicts sign epistasis (e.g. a mutation that is beneficial in a maladapted background may become deleterious in a well-adapted background due to optimum overshooting (Blanquart et al., 2014)). Overall, the success of FGM suggests that at least some of the repeatedly observed properties of the genotype–fitness relationship may be remarkably predictable without the need for any mechanistic knowledge.

4.3 | Genome-wide epistasis (intergenic epistasis)

The nature of epistasis appears to be less general across systems. Genome-wide deletion analyses (like the single gene and regulatory sequence studies) tend to find a predominance of negative interactions, though still alongside a substantial proportion of positive ones (Babu et al., 2014; Costanzo et al., 2010, 2016; Onge et al., 2007; Roguev et al., 2008; Szappanos et al., 2011), but it should be noted that these have mainly been performed in yeast for now. In line with flux balance analysis (FBA) pairwise gene perturbation predictions (He, Qian, Wang, Li, & Zhang, 2010; Segrè, DeLuna, Church, & Kishony, 2005; Szappanos et al., 2011), comparing interaction profiles for different genes has provided information on the topology

of their functional connections, providing "wiring diagrams" of cell function (Costanzo et al., 2016) (Figure 10). For example, genes from redundant pathways tend to undergo negative synergistic interactions with each other, and genes from the same pathway tend towards positive antagonistic interactions (perhaps largely responsible for the differential proportions of negative and positive epistasis) (Avery & Wasserman, 1992; Battle, Jonikas, Walter, Weissman, & Koller, 2010; Beltrao, Cagney, & Krogan, 2010; Breslow et al., 2008; Fu, Deng, Jin, Wang, & Yu, 2017; Lehner, 2011; Onge et al., 2007; Ye et al., 2005). As stated above, though, it is not clear that these rules should generalize to mutations other than the complete loss-of-function ones making up the vast majority of these data sets (hypomorphic alleles of essential genes are the exception) (Szathmary, 1993; Xu, Barker, & Gu, 2012).

On the other hand, epistasis between naturally occurring mutations in viral and cellular genomes appears to be biased towards positive interactions, with positive epistasis more common between the relatively frequent deleterious mutations and negative epistasis more common between rarely occurring beneficial mutations, that is an overall trend of antagonism (Bonhoeffer, 2004; Burch, 2004; Caudle et al., 2014; H.-H. Chou et al., 2011; Chou et al., 2014; de Visser et al., 2011; Flynn et al., 2013; Khan et al., 2011; Kryazhimskiy et al., 2014; J Lalić & Elena, 2012; Maisnier-Patin et al., 2005; Rokyta et al., 2011; Sanjuan, Moya, & Elena, 2004a; Schoustra, Hwang, Krug, & de Visser, 2016). This antagonism represents a kind of genomic buffering process: combinations of deleterious mutations are "less bad" than expected from simple additivity, and beneficial combinations are "less good", but a mechanistic explanation is lacking. Beneficial/deleterious combinations have rarely been studied explicitly. Further, the trend of diminishing returns between beneficial mutations found in single genes is often also found at the genome scale, suggesting analogously the saturation, and sometimes even optimum overshoot (Rokyta et al., 2011), of some phenotype contributing to fitness (Berger & Postma, 2014). These general fitness-level trends found for real mutations are rather encouraging for the predictability of adaptive dynamics, despite the underlying genetic and even phenotypic complexity (Kryazhimskiy et al., 2014).

5 | THE ENVIRONMENT

It is clear that the environment affects genotype–phenotype relationships, and so for a complete understanding of them we must consider them across different environments. Indeed, the few studies that explore large mutant sets across large numbers of environments find fundamental changes, such as the proportion of essential genes (Nichols et al., 2011). Even such ambitious large-scale studies explore only a vanishing fraction of potentially relevant fixed environments, though, not to mention dynamic ones.

Studies examining environmental effects therefore tend to have the ambition of proof of principle, rather than exhaustive sampling, and these have produced myriad examples of the environmental dependence of both mutation effects and epistasis. It
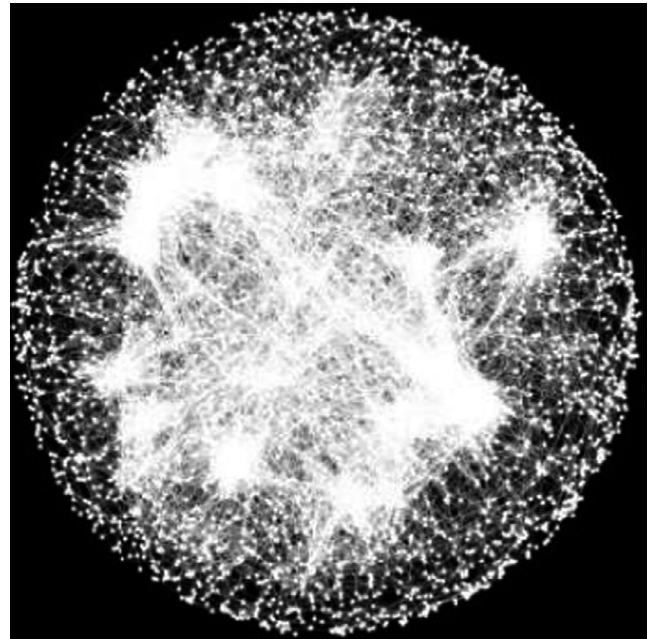
FIGURE 10 A genome-wide network of gene–gene interaction profile similarities. Nodes are yeast genes, and edges connect genes with similar genome-wide fitness interaction profiles, revealing functional modules. [From *Science*, *353*(6,306), aaf1420-2, A global genetic interaction network maps a wiring diagram of cellular function (Costanzo et al., 2016). Reprinted with permission from AAAS]

remains extremely difficult to form any general conclusions, as "environment" may refer to the concentration of small molecules (gene expression inducers, enzyme substrates, cofactors, antibiotics) which have some specific role in the system of study (Dean, 1995; Lagator et al., 2016; Lagator, Paixão, et al., 2017; Melnikov et al., 2014; Nghe et al., 2018; Shultzaberger et al., 2010; de Vos, Dawid, Sunderlikova, & Tans, 2015; de Vos, Poelwijk, Battich, Ndika, & Tans, 2013; Wrenbeck et al., 2017), precise physicochemical parameters known to matter in in vitro studies (Hayden, Ferrada, & Wagner, 2011; Hayden & Wagner, 2012), or more general "pleiotropic" factors such as temperature, chemical stresses, complex nutrients or even host organism (Bank et al., 2014; Caudle et al., 2014; Dandage et al., 2018; Flynn et al., 2013; Fragata et al., 2018; Hietpas, Bank, Jensen, & Bolon, 2013; Jagdishchandra Joshi & Prasad, 2014; Lalić, Cuevas, & Elena, 2011; Li & Zhang, 2018; Mavor et al., 2016). It will be important going forward to develop a more systematic approach to the environment, focussing either on environments likely to provide mechanistic insight into mutation effects, or simply on those most relevant to industry or nature.

6 | CONCLUDING REMARKS

Recent technological advances have now made it possible to score large-scale genetic libraries for a variety of phenotypes in massively parallel fashion ("deep mutational scanning") (Fowler &

Fields, 2014; Hietpas et al., 2011). In the short term, the resulting genotype–phenotype maps have illuminated several trends such as the bimodality of mutational effects and the pervasiveness of epistasis, and allowed for rigorous testing of mechanistic biological models. In the long term, they open the possibility of developing predictive genetic models, which would launch a new era in bioengineering, precision medicine, infectious disease control and bioconservation, to name a few. Such quantitative models are limited by the precision of the large data sets informing them (Rubin et al., 2017). Great strides have been made here—while early studies were limited in numbers of replicates due to their expense, genetic barcoding strategies now enable each genotype to be represented by multiple independent lineages in a single experiment, allowing characterization of biological and experimental noise and even error correction (Fowler et al., 2014; Mavor et al., 2016). Further, a critical decision for any deep-sequencing-based assay is how best to make use of a defined number of sequencing reads (*e.g.* depth per replicate, number of replicates, number of time points/selection cycles), and computational analysis has revealed some general guidelines for maximizing experimental power (Matuszewski, Hildebrandt, Ghenu, Jensen, & Bank, 2016).

An exciting future direction is the large-scale characterization of mutation effects at many phenotypic scales simultaneously (*e.g.* protein stability, protein activity, flux, expression, cell morphology, fitness), which could enable a direct and complete mechanistic description of the translation of genotype into high-level traits (Cambray et al., 2018). Indeed, certain phenotypes, such as metabolic flux (Sauer, 2004) and the set of –omes, have received very little direct attention, mostly due to the enormous technical challenges and cost involved in their high-throughput measurement/coupling to genotype libraries. One promising candidate, however, is the transcriptome, as it can be sequenced using RNA-seq: using the same emulsion-based technology that can enable distal genomic sites to be linked together for many single cells (Figure 9), the transcriptome can in principle be quantified for single cells while linking this information to the cells' genotypes with the use of unique cellular DNA barcodes (Adamson et al., 2016; Dixit et al., 2016). The transcriptomic impact of random mutations or a large set of transcriptional regulator mutations, for example, could thus be rapidly assessed and even linked to high-throughput fitness measurements.

Studies involving a few genes interacting at the level of elementary biological components (regulatory or metabolic motifs) can reveal surprisingly diverse types epistasis while allowing mechanistic interpretations of their origin. One promise is to integrate these local views to explain genome-wide epistasis from biological mechanisms. Smaller-scale studies may also allow the identification of generic principles which could be extrapolated to larger systems, for example the categories of phenotype to fitness relationships which generate certain types of epistasis or not. From a large-scale evolutionary perspective, it is not currently clear whether upscaling leads to an increase in the complexity of interactions, or could rather lead to a form of averaging, as suggested theoretically (Martin, 2014).

Finally, the systematic analysis of the effects of genome-wide combinations of point mutations still appears far out of reach, but a feasible next step might be introducing synthetic promoter libraries like those used in Keren et al. (2016) in front of *pairs* of genes across the genome and measuring the fitness effects. Although clearly artificial, and in some cases breaking regulatory links that are ensured by native promoters, such an experiment could provide quantitative two-gene expression–fitness landscapes for many pairs of genes, which should be an extremely important component of the genotype–fitness relationship, and which for now we are almost completely blind to (see Martin, 2016 for higher level 2-D trait–fitness landscapes in a multicellular organism).

With constantly improving technologies for reading and writing genotypes on a massive scale, and the application of experimental creativity, our mechanistic understanding of the genotype–phenotype relationship across different scales can only continue to flourish.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study.

ORCID

Harry Kemble https://orcid.org/0000-0003-1480-5873

Olivier Tenaillon https://orcid.org/0000-0002-3796-1601

REFERENCES

Adamson, B., Norman, T. M., Jost, M., Cho, M. Y., Nuñez, J. K., Chen, Y., … Weissman, J. S. (2016). A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell*, *167*(7), 1867–1882.e21. https://doi.org/10.1016/j.cell.2016.11.048

Agashe, D., Sane, M., Phalnikar, K., Diwan, G. D., Habibullah, A., Martinez-Gomez, N. C., … Marx, C. J. (2016). Large-effect beneficial synonymous mutations mediate rapid and parallel adaptation in a bacterium. *Molecular Biology and Evolution*, *33*(6), 1542–1553. https://doi.org/10.1093/molbev/msw035

Araya, C. L., Fowler, D. M., Chen, W., Muniez, I., Kelly, J. W., & Fields, S. (2012). A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proceedings of the National Academy of Sciences*, *109*(42), 16858–16863. https://doi.org/10.1073/pnas.1209751109

Avery, L., & Wasserman, S. (1992). Ordering gene function: The interpretation of epistasis in regulatory hierarchies. *Trends in Genetics*, *8*(9), 312–316. https://doi.org/10.1016/0168-9525(92)90263-4

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., ... Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Molecular Systems Biology*, 2, 2006.0008. https://doi.org/10.1038/msb4100050

Babu, M., Arnold, R., Bundalovic-Torma, C., Gagarinova, A., Wong, K. S., Kumar, A., ... Emili, A. (2014). Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in *Escherichia coli*. *PLoS Genetics*, 10(2), e1004120. https://doi.org/10.1371/journal.pgen.1004120

Badano, J. L., & Katsanis, N. (2002). Beyond Mendel: An evolving view of human genetic disease transmission. *Nature Reviews Genetics*, 3(10), 779–789. https://doi.org/10.1038/nrg910

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., ... Bulyk, M. L. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science*, 324(5935), 1720–1723. https://doi.org/10.1126/science.1162327

Baeza-Centurion, P., Miñana, B., Schmiedel, J. M., Valcárcel, J., & Lehner, B. (2019). Combinatorial genetics reveals a scaling law for the effects of mutations on splicing. *Cell*, 176(3), 549–563.e23. https://doi.org/10.1016/j.cell.2018.12.010

Bank, C., Hietpas, R. T., Jensen, J. D., & Bolon, D. N. A. (2015). A systematic survey of an intragenic epistatic landscape. *Molecular Biology and Evolution*, 32(1), 229–238. https://doi.org/10.1093/molbev/msu301

Bank, C., Hietpas, R. T., Wong, A., Bolon, D. N., & Jensen, J. D. (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(3), 841–852. https://doi.org/10.1534/genetics.113.156190

Bank, C., Matuszewski, S., Hietpas, R. T., & Jensen, J. D. (2016). On the (un)predictability of a large intragenic fitness landscape [Data set]. *Proceedings of the National Academy of Sciences*, 13: 14085–14090. https://doi.org/10.5061/dryad.th0rj

Barbieri, E. M., Muir, P., Akhuetie-Oni, B. O., Yellman, C. M., & Isaacs, F. J. (2017). Precise editing at DNA replication forks enables multiplex genome engineering in eukaryotes. *Cell*, 171(6), 1453–1467.e13. https://doi.org/10.1016/j.cell.2017.10.034

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., ... Myers, C. L. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature Methods*, 7(12), 1017–1024. https://doi.org/10.1038/nmeth.1534

Bataillon, T., & Bailey, S. F. (2014). Effects of new mutations on fitness: Insights from models and data: Effects of new mutations on fitness. *Annals of the New York Academy of Sciences*, 1320(1), 76–92. https://doi.org/10.1111/nyas.12460

Battle, A., Jonikas, M. C., Walter, P., Weissman, J. S., & Koller, D. (2010). Automated identification of pathways from quantitative genetic interaction data. *Molecular Systems Biology*, 6(379), https://doi.org/10.1038/msb.2010.27

Beltrao, P., Cagney, G., & Krogan, N. J. (2010). Quantitative genetic interactions reveal biological modularity. *Cell*, 141(5), 739–745. https://doi.org/10.1016/j.cell.2010.05.019

Bendixsen, D. P., Østman, B., & Hayden, E. J. (2017). Negative epistasis in experimental RNA fitness landscapes. *Journal of Molecular Evolution*, 85(5–6), 159–168. https://doi.org/10.1007/s00239-017-9817-5

Berger, D., & Postma, E. (2014). Biased estimates of diminishing-returns epistasis? *Empirical Evidence Revisited. Genetics*, 198(4), 1417–1420. https://doi.org/10.1534/genetics.114.169870

Bernet, G. P., & Elena, S. F. (2015). Distribution of mutational fitness effects and of epistasis in the 5' untranslated region of a plant RNA virus. *BMC Evolutionary Biology*, 15(1), https://doi.org/10.1186/s12862-015-0555-2

Bershtein, S., Mu, W., Serohijos, A. W. R., Zhou, J., & Shakhnovich, E. I. (2013). Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. *Molecular Cell*, 49(1), 133–144. https://doi.org/10.1016/j.molcel.2012.11.004

Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., & Tawfik, D. S. (2006). Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature*, 444(7121), 929–932. https://doi.org/10.1038/nature05385

Blanquart, F., Achaz, G., Bataillon, T., & Tenaillon, O. (2014). Properties of selected mutations and genotypic landscapes under Fisher's Geometric Model. *Evolution*, 68(12), 3537–3554. https://doi.org/10.1111/evo.12545

Bonhoeffer, S. (2004). Evidence for positive epistasis in HIV-1. *Science*, 306(5701), 1547–1550. https://doi.org/10.1126/science.1101786

Boutros, M., Kiger, A., Armknecht, S., Kerr, K., Hild, M., Koch, B., ... Perrimon, N. (2004). Genome-Wide RNAi analysis of growth and viability in *Drosophila* cells. *Science*, 303(5659), 832–835. https://doi.org/10.1126/science.1091266

Boyle, E. A., Andreasson, J. O. L., Chircus, L. M., Sternberg, S. H., Wu, M. J., Guegler, C. K., ... Greenleaf, W. J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proceedings of the National Academy of Sciences*, 114(21), 5461–5466. https://doi.org/10.1073/pnas.1700557114

Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C., & Kondrashov, F. A. (2012). Epistasis as the primary factor in molecular evolution. *Nature*, 490(7421), 535–538. https://doi.org/10.1038/nature11510

Breslow, D. K., Cameron, D. M., Collins, S. R., Schuldiner, M., Stewart-Ornstein, J., Newman, H. W., ... Weissman, J. S. (2008). A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nature Methods*, 5(8), 711–718. https://doi.org/10.1038/nmeth.1234

Buenrostro, J. D., Araya, C. L., Chircus, L. M., Layton, C. J., Chang, H. Y., Snyder, M. P., & Greenleaf, W. J. (2014). Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nature Biotechnology*, 32(6), 562–568. https://doi.org/10.1038/nbt.2880

Burch, C. L. (2004). Epistasis and Its Relationship to Canalization in the RNA Virus 6. *Genetics*, 167(2), 559–567. https://doi.org/10.1534/genetics.103.021196

Cambray, G., Guimaraes, J. C., & Arkin, A. P. (2018). Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. *Nature Biotechnology*, 36, 1005. https://doi.org/10.1038/nbt.4238

Carrasco, P., de la Iglesia, F., & Elena, S. F. (2007). Distribution of fitness and virulence effects caused by single-nucleotide substitutions in tobacco etch virus. *Journal of Virology*, 81(23), 12979–12984. https://doi.org/10.1128/JVI.00524-07

Caudle, S. B., Miller, C. R., & Rokyta, D. R. (2014). Environment determines epistatic patterns for a ssDNA virus. *Genetics*, 196(1), 267–279. https://doi.org/10.1534/genetics.113.158154

Chan, Y. H., Venev, S. V., Zeldovich, K. B., & Matthews, C. R. (2017). Correlation of fitness landscapes from three orthologous TIM barrels originates from sequence and structure constraints. *Nature Communications*, 8, 14614. https://doi.org/10.1038/ncomms14614

Chevin, L.-M., Martin, G., & Lenormand, T. (2010). Fisher's model and the genomics of adaptation: Restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution*, 64(11), 3213–3231. https://doi.org/10.1111/j.1558-5646.2010.01058.x

Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segre, D., & Marx, C. J. (2011). Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science*, 332(6034), 1190–1192. https://doi.org/10.1126/science.1203799

Chou, H.-H., Delaney, N. F., Draghi, J. A., & Marx, C. J. (2014). Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLoS Genetics*, 10(2), e1004149. https://doi.org/10.1371/journal.pgen.1004149

Collins, S. R., Miller, K. M., Maas, N. L., Roguev, A., Fillingham, J., Chu, C. S., ... Krogan, N. J. (2007). Functional dissection of protein complexes

involved in yeast chromosome biology using a genetic interaction map. *Nature*, *446*(7137), 806–810. https://doi.org/10.1038/nature05649

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., … Boone, C. (2010). The genetic landscape of a cell. *Science*, *327*(5964), 425–431. https://doi.org/10.1126/science.1180823

Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., … Boone, C. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science*, *353*(6306), aaf1420–aaf1420. https://doi.org/10.1126/science.aaf1420

Cuevas, J. M., Domingo-Calap, P., & Sanjuán, R. (2012). The fitness effects of synonymous mutations in DNA and RNA viruses. *Molecular Biology and Evolution*, *29*(1), 17–20. https://doi.org/10.1093/molbev/msr179

Dandage, R., Pandey, R., Jayaraj, G., Rai, M., Berger, D., & Chakraborty, K. (2018). Differential strengths of molecular determinants guide environment specific mutational fates. *PLOS Genetics*, *14*(5), e1007419. https://doi.org/10.1371/journal.pgen.1007419

Davierwala, A. P., Haynes, J., Li, Z., Brost, R. L., Robinson, M. D., Yu, L., … Hughes, T. R. (2005). The synthetic genetic interaction spectrum of essential genes. *Nature Genetics*, *37*(10), 1147–1152. https://doi.org/10.1038/ng1640

de Visser, J. A. G. M., Cooper, T. F., & Elena, S. F. (2011). The causes of epistasis. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1725), 3617–3624. https://doi.org/10.1098/rspb.2011.1537

de Vos, M. G. J., Dawid, A., Sunderlikova, V., & Tans, S. J. (2015). Breaking evolutionary constraint with a tradeoff ratchet. *Proceedings of the National Academy of Sciences*, *112*(48), 14906–14911. https://doi.org/10.1073/pnas.1510282112

de Vos, M. G. J., Poelwijk, F. J., Battich, N., Ndika, J. D. T., & Tans, S. J. (2013). Environmental dependence of genetic constraint. *PLoS Genetics*, *9*(6), e1003580. https://doi.org/10.1371/journal.pgen.1003580

Dean, A. M. (1995). A molecular investigation of genotype by environment interactions. *Genetics*, *139*, 19–33.

Decourty, L., Saveanu, C., Zemam, K., Hantraye, F., Frachon, E., Rousselle, J.-C., … Jacquier, A. (2008). Linking functionally related genes by sensitive and quantitative characterization of genetic interaction profiles. *Proceedings of the National Academy of Sciences*, *105*(15), 5821–5826. https://doi.org/10.1073/pnas.0710533105

Dekel, E., & Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature*, *436*(7050), 588–592. https://doi.org/10.1038/nature03842

DePristo, M. A., Weinreich, D. M., & Hartl, D. L. (2005). Missense meanderings in sequence space: A biophysical view of protein evolution. *Nature Reviews Genetics*, *6*(9), 678–687. https://doi.org/10.1038/nrg1672

Dipple, K. M., & McCabe, E. R. B. (2000). Modifier genes convert "Simple" mendelian disorders to complex traits. *Molecular Genetics and Metabolism*, *71*(1–2), 43–50. https://doi.org/10.1006/mgme.2000.3052

Diss, G., & Lehner, B. (2018). The genetic landscape of a physical interaction. *ELife*, *7*, e32472. https://doi.org/10.7554/elife.32472.001

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., … Regev, A. (2016). Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*, *167*(7), 1853–1866.e17. https://doi.org/10.1016/j.cell.2016.11.038

Domingo, J., Diss, G., & Lehner, B. (2018). Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature*, *558*(7708), 117–121. https://doi.org/10.1038/s41586-018-0170-7

Domingo-Calap, P., Cuevas, J. M., & Sanjuán, R. (2009). The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genetics*, *5*(11), e1000742. https://doi.org/10.1371/journal.pgen.1000742

Douglas, A. C., Smith, A. M., Sharifpoor, S., Yan, Z., Durbic, T., Heisler, L. E., … Andrews, B. J. (2012). Functional analysis with a barcoder yeast.

Gene Overexpression System. G3: Genes|genomes|genetics, *2*(10), 1279–1289. https://doi.org/10.1534/g3.112.003400

Drummond, D. A., & Wilke, C. O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*, *134*(2), 341–352. https://doi.org/10.1016/j.cell.2008.05.042

Dykhuizen, D. E., Dean, A. M., & Hartl, D. L. (1987). Metabolic flux and fitness. *Genetics*, *115*, 25–31.

Echave, J., & Wilke, C. O. (2017). Biophysical models of protein evolution: understanding the patterns of evolutionary sequence divergence. *Annual Review of Biophysics*, *46*(1), 85–103. https://doi.org/10.1146/annurev-biophys-070816-033819

Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, *8*(8), 610–618. https://doi.org/10.1038/nrg2146

Firnberg, E., Labonte, J. W., Gray, J. J., & Ostermeier, M. (2014). A comprehensive, high-resolution map of a gene's fitness landscape. *Molecular Biology and Evolution*, *31*(6), 1581–1592. https://doi.org/10.1093/molbev/msu081

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford, UK: Oxford University Press.

Flynn, K. M., Cooper, T. F., Moore, F.-B.-G., & Cooper, V. S. (2013). The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. *PLoS Genetics*, *9*(4), e1003426. https://doi.org/10.1371/journal.pgen.1003426

Fowler, D. M., Araya, C. L., Fleishman, S. J., Kellogg, E. H., Stephany, J. J., Baker, D., & Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. *Nature Methods*, *7*(9), 741–746. https://doi.org/10.1038/nmeth.1492

Fowler, D. M., & Fields, S. (2014). Deep mutational scanning: A new style of protein science. *Nature Methods*, *11*(8), 801–807. https://doi.org/10.1038/nmeth.3027

Fowler, D. M., Stephany, J. J., & Fields, S. (2014). Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nature Protocols*, *9*(9), 2267–2284. https://doi.org/10.1038/nprot.2014.153

Fragata, I., Matuszewski, S., Schmitz, M. A., Bataillon, T., Jensen, J. D., & Bank, C. (2018). The fitness landscape of the codon space across environments. *Heredity*, https://doi.org/10.1038/s41437-018-0125-7

Fu, C., Deng, S., Jin, G., Wang, X., & Yu, Z.-G. (2017). Bayesian network model for identification of pathways by integrating protein interaction with genetic interaction data. *BMC Systems Biology*, *11*(S4), https://doi.org/10.1186/s12918-017-0454-9

Fuchs, F., Pau, G., Kranz, D., Sklyar, O., Budjan, C., Steinbrink, S., … Boutros, M. (2010). Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. *Molecular Systems Biology*, *6*, https://doi.org/10.1038/msb.2010.25

Gagarinova, A., Stewart, G., Samanfar, B., Phanse, S., White, C. A., Aoki, H., … Emili, A. (2016). Systematic genetic screens reveal the dynamic global functional organization of the bacterial translation machinery. *Cell Reports*, *17*(3), 904–916. https://doi.org/10.1016/j.celrep.2016.09.040

Geertz, M., Shore, D., & Maerkl, S. J. (2012). Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proceedings of the National Academy of Sciences*, *109*(41), 16540–16545. https://doi.org/10.1073/pnas.1206011109

Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balázsi, G., Ravasz, E., Daugherty, M. D., … Osterman, A. L. (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *Journal of Bacteriology*, *185*(19), 5673–5684. https://doi.org/10.1128/JB.185.19.5673-5684.2003

Giaever, G., Chu, A. M., Ni, L. I., Connelly, C., Riles, L., Véronneau, S., … Johnston, M. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, *418*(6896), 387–391. https://doi.org/10.1038/nature00935

Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., … Venter, J. C. (2006). Essential genes of a minimal

bacterium. *Proceedings of the National Academy of Sciences*, *103*(2), 425–430. https://doi.org/10.1073/pnas.0510013103

Guenther, U.-P., Yandek, L. E., Niland, C. N., Campbell, F. E., Anderson, D., Anderson, V. E., … Jankowsky, E. (2013). Hidden specificity in an apparently nonspecific RNA-binding protein. *Nature*, *502*(7471), 385–388. https://doi.org/10.1038/nature12543

Haimovich, A. D., Muir, P., & Isaacs, F. J. (2015). Genomes by design. *Nature Reviews Genetics*, *16*(9), 501–516. https://doi.org/10.1038/nrg3956

Haliburton, J. R., Shao, W., Deutschbauer, A., Arkin, A., & Abate, A. R. (2017). Genetic interaction mapping with microfluidic-based single cell sequencing. *PLoS ONE*, *12*(2), e0171302. https://doi.org/10.1371/journal.pone.0171302

Hansen, T. F. (2013). Why epistasis is important for selection and adaptation: Perspective. *Evolution*, *67*(12), 3501–3511. https://doi.org/10.1111/evo.12214

Hayden, E. J., Ferrada, E., & Wagner, A. (2011). Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature*, *474*(7349), 92–95. https://doi.org/10.1038/nature10083

Hayden, E. J., & Wagner, A. (2012). Environmental change exposes beneficial epistatic interactions in a catalytic RNA. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1742), 3418–3425. https://doi.org/10.1098/rspb.2012.0956

He, X., Qian, W., Wang, Z., Li, Y., & Zhang, J. (2010). Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. *Nature Genetics*, *42*(3), 272–276. https://doi.org/10.1038/ng.524

Hietpas, R. T., Bank, C., Jensen, J. D., & Bolon, D. N. A. (2013). Shifting fitness landscapes in response to altered environments. *Evolution*, *67*(12), 3512–3522. https://doi.org/10.1111/evo.12207

Hietpas, R. T., Jensen, J. D., & Bolon, D. N. A. (2011). Experimental illumination of a fitness landscape. *Proceedings of the National Academy of Sciences*, *108*(19), 7896–7901. https://doi.org/10.1073/pnas.1016024108

Hutchison, C. A. III, Peterson, S., Gill, S., Cline, R., White, O., Fraser, C., … Craig, J., V. (1999). Global transposon mutagenesis and a minimal mycoplasma genome. *Science*, *286*(5447), 2165–2169. https://doi.org/10.1126/science.286.5447.2165

Jacquier, H., Birgy, A., Le Nagard, H., Mechulam, Y., Schmitt, E., Glodt, J., … Tenaillon, O. (2013). Capturing the mutational landscape of the beta-lactamase TEM-1. *Proceedings of the National Academy of Sciences*, *110*(32), 13067–13072. https://doi.org/10.1073/pnas.1215206110

Jaffe, M., Sherlock, G., & Levy, S. F. (2017). iSeq: A new double-barcode method for detecting dynamic genetic interactions in yeast. *G3: Genes*, *Genomes*, *Genetics*, *7*(1), 143–153. https://doi.org/10.1534/g3.116.034207

Jagdishchandra Joshi, C., & Prasad, A. (2014). Epistatic interactions among metabolic genes depend upon environmental conditions. *Molecular BioSystems*, *10*(10), 2578–2589. https://doi.org/10.1039/C4MB00181H

Jiang, L. I., Liu, P., Bank, C., Renzette, N., Prachanronarong, K., Yilmaz, L. S., … Bolon, D. N. A. (2016). A balance between inhibitor binding and substrate processing confers influenza drug resistance. *Journal of Molecular Biology*, *428*(3), 538–553. https://doi.org/10.1016/j.jmb.2015.11.027

Jiang, L., Mishra, P., Hietpas, R. T., Zeldovich, K. B., & Bolon, D. N. A. (2013). Latent effects of Hsp90 mutants revealed at reduced expression levels. *PLoS Genetics*, *9*(6), e1003600. https://doi.org/10.1371/journal.pgen.1003600

Kacser, H., & Burns, J. A. (1981). The molecular basis of dominance. *Genetics*, *97*, 639–666.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., … Ahringer, J. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, *421*(6920), 231–237. https://doi.org/10.1038/nature01278

Keren, L., Hausser, J., Lotan-Pompan, M., Vainberg Slutskin, I., Alisar, H., Kaminski, S., … Segal, E. (2016). Massively parallel interrogation of the effects of gene expression levels on fitness. *Cell*, *166*(5), 1282–1294.e18. https://doi.org/10.1016/j.cell.2016.07.024

Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E., & Cooper, T. F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, *332*(6034), 1193–1196. https://doi.org/10.1126/science.1203801

Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., … Hoe, K.-L. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. *Nature Biotechnology*, *28*(6), 617–623. https://doi.org/10.1038/nbt.1628

Kim, I., Miller, C. R., Young, D. L., & Fields, S. (2013). High-throughput analysis of *in vivo* protein stability. *Molecular & Cellular Proteomics*, *12*(11), 3370–3378. https://doi.org/10.1074/mcp.O113.031708

Kimura, M., & Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics*, *54*(6), 1337–1351.

Kinney, J. B., Murugan, A., Callan, C. G., & Cox, E. C. (2010). Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proceedings of the National Academy of Sciences*, *107*(20), 9158–9163. https://doi.org/10.1073/pnas.1004290107

Klesmith, J. R., Bacik, J.-P., Michalczyk, R., & Whitehead, T. A. (2015). Comprehensive sequence-flux mapping of a levoglucosan utilization pathway in *E. coli*. *ACS Synthetic Biology*, *4*(11), 1235–1243. https://doi.org/10.1021/acssynbio.5b00131

Kobori, S., & Yokobayashi, Y. (2016). High-throughput mutational analysis of a twister ribozyme. *Angewandte Chemie International Edition*, *55*(35), 10354–10357. https://doi.org/10.1002/anie.201605470

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*, *336*, 435. https://doi.org/10.1038/336435a0

Kondrashov, F. A., & Kondrashov, A. S. (2001). Multidimensional epistasis and the disadvantage of sex. *Proceedings of the National Academy of Sciences*, *98*(21), 12089–12092. https://doi.org/10.1073/pnas.211214298

Kryazhimskiy, S., Rice, D. P., Jerison, E. R., & Desai, M. M. (2014). Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, *344*(6191), 1519–1522. https://doi.org/10.1126/science.1250939

Kunkel, T. A., Bebenek, K., & McClary, J. (1991). Efficient site-directed mutagenesis using uracil-containing DNA. *Methods in Enzymology*, *204*, 125–139.

Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., … Myers, C. L. (2018). Systematic analysis of complex genetic interactions. *Science*, *360*(6386), eaao1729. https://doi.org/10.1126/science.aao1729

Kvitek, D. J., & Sherlock, G. (2011). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genetics*, *7*(4), e1002056. https://doi.org/10.1371/journal.pgen.1002056

Kwasnieski, J. C., Mogno, I., Myers, C. A., Corbo, J. C., & Cohen, B. A. (2012). Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proceedings of the National Academy of Sciences*, *109*(47), 19498–19503. https://doi.org/10.1073/pnas.1210678109

Lagator, M., Igler, C., Moreno, A. B., Guet, C. C., & Bollback, J. P. (2016). Epistatic interactions in the Arabinose *Cis* -regulatory element. *Molecular Biology and Evolution*, *33*(3), 761–769. https://doi.org/10.1093/molbev/msv269

Lagator, M., Paixão, T., Barton, N. H., Bollback, J. P., & Guet, C. C. (2017). On the mechanistic nature of epistasis in a canonical cis-regulatory element. *eLife*, *6*, e25192. https://doi.org/10.7554/eLife.25192

Lagator, M., Sarikas, S., Acar, H., Bollback, J. P., & Guet, C. C. (2017). Regulatory network structure determines patterns of intermolecular epistasis. *eLife*, *6*, e28921. https://doi.org/10.7554/eLife.28921

Lalić, J., Cuevas, J. M., & Elena, S. F. (2011). Effect of host species on the distribution of mutational fitness effects for an RNA virus. *PLoS Genetics*, 7(11), e1002378. https://doi.org/10.1371/journal.pgen.1002378

Lalić, J., & Elena, S. F. (2012). Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. *Heredity*, 109(2), 71–77. https://doi.org/10.1038/hdy.2012.15

Layton, C. J., McMahon, P. L., & Greenleaf, W. J. (2019). Large-scale, quantitative protein assays on a high-throughput dna sequencing chip. *Molecular Cell*, 73(1), 1075-1082. https://doi.org/10.1016/j.molcel.2019.02.019

Lehner, B. (2011). Molecular mechanisms of epistasis within and between genes. *Trends in Genetics*, 27(8), 323–331. https://doi.org/10.1016/j.tig.2011.05.007

Lehner, B. (2013). Genotype to phenotype: Lessons from model organisms for human genetics. *Nature Reviews Genetics*, 14(3), 168–178. https://doi.org/10.1038/nrg3404

Levin, A. M., & Weiss, G. A. (2006). Optimizing the affinity and specificity of proteins with molecular display. *Molecular BioSystems*, 2(1), 49–57. https://doi.org/10.1039/B511782H

Levine, M., Cattoglio, C., & Tjian, R. (2014). Looping back to leap forward: transcription enters a New Era. *Cell*, 157(1), 13–25. https://doi.org/10.1016/j.cell.2014.02.009

Levo, M., & Segal, E. (2014). In pursuit of design principles of regulatory sequences. *Nature Reviews Genetics*, 15(7), 453–468. https://doi.org/10.1038/nrg3684

Li, C., Qian, W., Maclean, C. J., & Zhang, J. (2016). The fitness landscape of a tRNA gene. *Science*, 352(6287), 837–840. https://doi.org/10.1126/science.aae0568

Li, C., & Zhang, J. (2018). Multi-environment fitness landscapes of a tRNA gene. *Nature Ecology & Evolution*, 2(6), 1025–1032. https://doi.org/10.1038/s41559-018-0549-8

Lunzer, M., Miller, S. P., Felsheim, R., & Dean, A. M. (2005). The biochemical architecture of an ancient adaptive landscape. *Science*, 310(5747), 499–501. https://doi.org/10.1126/science.1115649

MacLean, R. C., Perron, G. G., & Gardner, A. (2010). Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics*, 186(4), 1345–1354. https://doi.org/10.1534/genetics.110.123083

Maerkl, S. J., & Quake, S. R. (2009). Experimental determination of the evolvability of a transcription factor. *Proceedings of the National Academy of Sciences*, 106(44), 18650–18655. https://doi.org/10.1073/pnas.0907688106

Maisnier-Patin, S., Roth, J. R., Fredriksson, Å., Nyström, T., Berg, O. G., & Andersson, D. I. (2005). Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nature Genetics*, 37(12), 1376–1379. https://doi.org/10.1038/ng1676

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747–753. https://doi.org/10.1038/nature08494

Martin, C. H. (2016). Context dependence in complex adaptive landscapes: Frequency and trait-dependent selection surfaces within an adaptive radiation of Caribbean pupfishes: Context-dependent fitness surfaces. *Evolution*, 70(6), 1265–1282. https://doi.org/10.1111/evo.12932

Martin, G. (2014). Fisher's geometrical model emerges as a property of complex integrated phenotypic networks. *Genetics*, 197(1), 237–255. https://doi.org/10.1534/genetics.113.160325

Martin, G., Elena, S. F., & Lenormand, T. (2007). Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature Genetics*, 39(4), 555–560. https://doi.org/10.1038/ng1998

Martin, G., & Lenormand, T. (2006). A general multivariate extension of fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution*, 60(5), 893–907. https://doi.org/10.1111/j.0014-3820.2006.tb01169.x

Matuszewski, S., Hildebrandt, M. E., Ghenu, A.-H., Jensen, J. D., & Bank, C. (2016). A statistical guide to the design of deep mutational scanning experiments. *Genetics*, 204(1), 77–87. https://doi.org/10.1534/genetics.116.190462

Mavor, D., Barlow, K., Thompson, S., Barad, B. A., Bonny, A. R., Cario, C. L., … Fraser, J. S. (2016). Determination of ubiquitin fitness landscapes under different chemical stresses in a classroom setting. *eLife*, 5, e15802. https://doi.org/10.7554/eLife.15802

McLaughlin, R. N. Jr, Poelwijk, F. J., Raman, A., Gosal, W. S., & Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. *Nature*, 491(7422), 138–142. https://doi.org/10.1038/nature11500

Melamed, D., Young, D. L., Gamble, C. E., Miller, C. R., & Fields, S. (2013). Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein. *RNA*, 19(11), 1537–1551. https://doi.org/10.1261/rna.040709.113

Melnikov, A., Rogov, P., Wang, L., Gnirke, A., & Mikkelsen, T. S. (2014). Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Research*, 42(14), e112. https://doi.org/10.1093/nar/gku511

Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D. J., Salichos, L., … Gerstein, M. (2016). The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biology*, 17(1), https://doi.org/10.1186/s13059-016-0917-0

Mustonen, V., Kinney, J., Callan, C. G., & Lassig, M. (2008). Energy-dependent fitness: A quantitative model for the evolution of yeast transcription factor binding sites. *Proceedings of the National Academy of Sciences*, 105(34), 12376–12381. https://doi.org/10.1073/pnas.0805909105

Nagel, A. C., Joyce, P., Wichman, H. A., & Miller, C. R. (2012). Stickbreaking: A novel fitness landscape model that harbors epistasis and is consistent with commonly observed patterns of adaptive evolution. *Genetics*, 190(2), 655–667. https://doi.org/10.1534/genetics.111.132134

Nghe, P., Kogenaru, M., & Tans, S. J. (2018). Sign epistasis caused by hierarchy within signalling cascades. *Nature Communications*, 9(1), 1451. https://doi.org/10.1038/s41467-018-03644-8

Nichols, R. J., Sen, S., Choo, Y. J., Beltrao, P., Zietek, M., Chaba, R., … Gross, C. A. (2011). Phenotypic landscape of a bacterial cell. *Cell*, 144(1), 143–156. https://doi.org/10.1016/j.cell.2010.11.052

Niederberger, P., Prasad, R., Miozzari, G., & Kacser, H. (1992). A strategy for increasing an in vivo flux by genetic manipulations. The tryptophan system of yeast. *Biochemical Journal*, 287(2), 473–479. https://doi.org/10.1042/bj2870473

Noderer, W. L., Flockhart, R. J., Bhaduri, A., Diaz de Arce, A. J., Zhang, J., Khavari, P. A., & Wang, C. L. (2014). Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Molecular Systems Biology*, 10(8), 748–748. https://doi.org/10.15252/msb.20145136

Nutiu, R., Friedman, R. C., Luo, S., Khrebtukova, I., Silva, D., Li, R., … Burge, C. B. (2011). Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nature Biotechnology*, 29(7), 659–664. https://doi.org/10.1038/nbt.1882

Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature*, 246, 96–98. https://doi.org/10.1038/246096a0

Olson, C. A., Wu, N. C., & Sun, R. (2014). A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Current Biology*, 24(22), 2643–2651. https://doi.org/10.1016/j.cub.2014.09.072

Onge, R. P. S., Mani, R., Oh, J., Proctor, M., Fung, E., Davis, R. W., … Giaever, G. (2007). Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nature Genetics*, 39(2), 199–206. https://doi.org/10.1038/ng1948

Otwinowski, J. (2018). Biophysical inference of epistasis and the effects of mutations on protein stability and function. *Molecular Biology*

and Evolution, 35(10), 2345–2354. https://doi.org/10.1093/molbev/msy141

Otwinowski, J., McCandlish, D. M., & Plotkin, J. B. (2018). Inferring the shape of global epistasis. *Proceedings of the National Academy of Sciences*, 115(32), E7550–E7558. https://doi.org/10.1073/pnas.1804015115

Pál, G., Kouadio, J.-L.-K., Artis, D. R., Kossiakoff, A. A., & Sidhu, S. S. (2006). Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. *Journal of Biological Chemistry*, 281(31), 22378–22385. https://doi.org/10.1074/jbc.M603826200

Patel, L., Abate, C., & Curran, T. (1990). Altered protein conformation on DNA binding by Fos and Jun. *Nature*, 347, 572–575. https://doi.org/10.1038/347572a0

Perfeito, L., Ghozzi, S., Berg, J., Schnetz, K., & Lässig, M. (2011). Nonlinear fitness landscape of a molecular pathway. *PLoS Genetics*, 7(7), e1002160. https://doi.org/10.1371/journal.pgen.1002160

Peris, J. B., Davis, P., Cuevas, J. M., Nebot, M. R., & Sanjuan, R. (2010). Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. *Genetics*, 185(2), 603–609. https://doi.org/10.1534/genetics.110.115162

Phillips, P. C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9(11), 855–867. https://doi.org/10.1038/nrg2452

Puchta, O., Cseke, B., Czaja, H., Tollervey, D., Sanguinetti, G., & Kudla, G. (2016). Network of epistatic interactions within a yeast snoRNA. *Science*, 352(6287), 840–844. https://doi.org/10.1126/science.aaf0965

Rajkumar, A. S., Dénervaud, N., & Maerkl, S. J. (2013). Mapping the fine structure of a eukaryotic promoter input-output function. *Nature Genetics*, 45(10), 1207–1215. https://doi.org/10.1038/ng.2729

Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S. R., … Krogan, N. J. (2008). Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*, 322(5900), 405–410. https://doi.org/10.1126/science.1162609

Rojas Echenique, J. I., Kryazhimskiy, S., Nguyen Ba, A. N., & Desai, M. M. (2019). Modular epistasis and the compensatory evolution of gene deletion mutants. *PLOS Genetics*, 15(2), e1007958. https://doi.org/10.1371/journal.pgen.1007958

Rokyta, D. R., Joyce, P., Caudle, S. B., Miller, C., Beisel, C. J., & Wichman, H. A. (2011). Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. *PLoS Genetics*, 7(6), e1002075. https://doi.org/10.1371/journal.pgen.1002075

Roscoe, B. P., Thayer, K. M., Zeldovich, K. B., Fushman, D., & Bolon, D. N. A. (2013). Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *Journal of Molecular Biology*, 425(8), 1363–1377. https://doi.org/10.1016/j.jmb.2013.01.032

Roy, K. R., Smith, J. D., Vonesch, S. C., Lin, G., Tu, C. S., Lederer, A. R., … Steinmetz, L. M. (2018). Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nature Biotechnology*, 36(6), 512–520. https://doi.org/10.1038/nbt.4137

Rubin, A. F., Gelman, H., Lucas, N., Bajjalieh, S. M., Papenfuss, A. T., Speed, T. P., & Fowler, D. M. (2017). A statistical framework for analyzing deep mutational scanning data. *Genome Biology*, 18(1), 150. https://doi.org/10.1186/s13059-017-1272-5

Sameith, K., Amini, S., Groot Koerkamp, M. J. A., van Leenen, D., Brok, M., Brabers, N., … Kemmeren, P. (2015). A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions. *BMC Biology*, 13(1), https://doi.org/10.1186/s12915-015-0222-5

Sanjuan, R., Moya, A., & Elena, S. F. (2004a). The contribution of epistasis to the architecture of fitness in an RNA virus. *Proceedings of the National Academy of Sciences*, 101(43), 15376–15379. https://doi.org/10.1073/pnas.0404125101

Sanjuan, R., Moya, A., & Elena, S. F. (2004b). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proceedings of the National Academy of Sciences*, 101(22), 8396–8401. https://doi.org/10.1073/pnas.0400146101

Sarkisyan, K. S., Bolotin, D. A., Meer, M. V., Usmanova, D. R., Mishin, A. S., Sharonov, G. V., … Kondrashov, F. A. (2016). Local fitness landscape of the green fluorescent protein. *Nature*, 533(7603), 397–401. https://doi.org/10.1038/nature17995

Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis: Genes required for mycobacterial growth. *Molecular Microbiology*, 48(1), 77–84. https://doi.org/10.1046/j.1365-2958.2003.03425.x

Sauer, U. (2004). High-throughput phenomics: Experimental methods for mapping fluxomes. *Current Opinion in Biotechnology*, 15(1), 58–63. https://doi.org/10.1016/j.copbio.2003.11.001

Schenk, M. F., Szendro, I. G., Salverda, M. L. M., Krug, J., & de Visser, J. A. G. M. (2013). Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. *Molecular Biology and Evolution*, 30(8), 1779–1787. https://doi.org/10.1093/molbev/mst096

Schoustra, S., Hwang, S., Krug, J., & de Visser, J. A. G. M. (2016). Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. *Proceedings of the Royal Society B: Biological Sciences*, 283(1837), 20161376. https://doi.org/10.1098/rspb.2016.1376

Schuldiner, M., Collins, S. R., Thompson, N. J., Denic, V., Bhamidipati, A., Punna, T., … Krogan, N. J. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell*, 123(3), 507–519. https://doi.org/10.1016/j.cell.2005.08.031

Scott, J., & Smith, G. (1990). Searching for peptide ligands with an epitope library. *Science*, 249(4967), 386–390. https://doi.org/10.1126/science.1696028

Scriver, C. R., & Waters, P. J. (1999). Monogenic traits are not simple: Lessons from phenylketonuria. *Trends in Genetics*, 15(7), 267–272. https://doi.org/10.1016/S0168-9525(99)01761-8

Segrè, D., DeLuna, A., Church, G. M., & Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nature Genetics*, 37(1), 77–83. https://doi.org/10.1038/ng1489

Serohijos, A. W. R., Rimas, Z., & Shakhnovich, E. I. (2012). Protein biophysics explains why highly abundant proteins evolve slowly. *Cell Reports*, 2(2), 249–256. https://doi.org/10.1016/j.celrep.2012.06.022

Serohijos, A. W. R., & Shakhnovich, E. I. (2014). Contribution of selection for protein folding stability in shaping the patterns of polymorphisms in coding regions. *Molecular Biology and Evolution*, 31(1), 165–176. https://doi.org/10.1093/molbev/mst189

Shah, P., McCandlish, D. M., & Plotkin, J. B. (2015). Contingency and entrenchment in protein evolution under purifying selection. *Proceedings of the National Academy of Sciences*, 112(25), E3226–E3235. https://doi.org/10.1073/pnas.1412933112

Shultzaberger, R. K., Maerkl, S. J., Kirsch, J. F., & Eisen, M. B. (2012). Probing the informational and regulatory plasticity of a transcription factor DNA–Binding domain. *PLoS Genetics*, 8(3), e1002614. https://doi.org/10.1371/journal.pgen.1002614

Shultzaberger, R. K., Malashock, D. S., Kirsch, J. F., & Eisen, M. B. (2010). The fitness landscapes of cis-acting binding sites in different promoter and environmental contexts. *PLoS Genetics*, 6(7), e1001042. https://doi.org/10.1371/journal.pgen.1001042

Sikosek, T., & Chan, H. S. (2014). Biophysics of protein evolution and evolutionary protein biophysics. *Journal of the Royal Society Interface*, 11(100), 20140419. https://doi.org/10.1098/rsif.2014.0419

Smith, G. (1985). Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science*, 228(4705), 1315–1317. https://doi.org/10.1126/science.4001944

Starita, L. M., Pruneda, J. N., Lo, R. S., Fowler, D. M., Kim, H. J., Hiatt, J. B., … Klevit, R. E. (2013). Activity-enhancing mutations in an E3 ubiquitin

ligase identified by high-throughput mutagenesis. *Proceedings of the National Academy of Sciences*, 110(14), E1263–E1272. https://doi.org/10.1073/pnas.1303309110

Starr, T. N., & Thornton, J. W. (2016). Epistasis in protein evolution. *Protein Science: A Publication of the Protein Society*, 25(7), 1204–1218. https://doi.org/10.1002/pro.2897

Szafraniec, K., Wloch, D. M., Sliwa, P., Borts, R. H., & Korona, R. (2003). Small fitness effects and weak genetic interactions between deleterious mutations in heterozygous loci of the yeast Saccharomyces cerevisiae. *Genetical Research*, 82(1), 19–31. https://doi.org/10.1017/S001667230300630X

Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., ... Papp, B. (2011). An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nature Genetics*, 43(7), 656–662. https://doi.org/10.1038/ng.846

Szathmary, E. (1993). Do deleterious mutations act synergistically? Metabolic Control Theory provides a partial answer. *Genetics*, 133, 127–132.

Tenaillon, O. (2014). The utility of fisher's geometric model in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics*, 45(1), 179–201. https://doi.org/10.1146/annurev-ecolsys-120213-091846

Thompson, K. S., Vinson, C. R., & Freire, E. (1993). Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4. *Biochemistry*, 32(21), 5491–5496. https://doi.org/10.1021/bi00072a001

Tischler, J., Lehner, B., Chen, N., & Fraser, A. G. (2006). Combinatorial RNA interference in Caenorhabditis elegans reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. *Genome Biology*, 7, R69.

Todeschini, A.-L., Georges, A., & Veitia, R. A. (2014). Transcription factors: Specific DNA binding and specific gene regulation. *Trends in Genetics*, 30(6), 211–219. https://doi.org/10.1016/j.tig.2014.04.002

Tokuriki, N., & Tawfik, D. S. (2009). Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology*, 19(5), 596–604. https://doi.org/10.1016/j.sbi.2009.08.003

Tong, A. H. Y. (2004). Global mapping of the yeast genetic interaction network. *Science*, 303(5659), 808–813. https://doi.org/10.1126/science.1091317

Towbin, B. D., Korem, Y., Bren, A., Doron, S., Sorek, R., & Alon, U. (2017). Optimality and sub-optimality in a bacterial growth law. *Nature Communications*, 8, 14123. https://doi.org/10.1038/ncomms14123

Trindade, S., Sousa, A., & Gordo, I. (2012). Antibiotic resistance and stress in the light of Fisher's model. *Evolution*, 66(12), 3815–3824. https://doi.org/10.1111/j.1558-5646.2012.01722.x

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., ... Hahn, W. C. (2017). Defining a cancer dependency map. *Cell*, 170(3), 564–576.e16. https://doi.org/10.1016/j.cell.2017.06.010

Ursell, T., Lee, T. K., Shiomi, D., Shi, H., Tropini, C., Monds, R. D., ... Huang, K. C. (2017). Rapid, precise quantification of bacterial cellular dimensions across a genomic-scale knockout library. *BMC Biology*, 15(1), https://doi.org/10.1186/s12915-017-0348-8

van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq: High-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, 6(10), 767–772. https://doi.org/10.1038/nmeth.1377

van Opijnen, T., Lazinski, D. W., & Camilli, A. (2014). Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms: Tn-seq: High-Throughput Sequencing for Microorganisms. In F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, & K. Struhl (Eds.), *Current Protocols in Molecular Biology* (pp. 7.16.1–7.16.24). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/0471142727.mb0716s106

Venkataram, S., Dunn, B., Li, Y., Agarwala, A., Chang, J., Ebel, E. R., ... Petrov, D. A. (2016). Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast. *Cell*, 166(6), 1585–1596.e22. https://doi.org/10.1016/j.cell.2016.08.002

Vilar, J. M. G. (2010). Accurate prediction of gene expression by integration of DNA sequence statistics with detailed modeling of transcription regulation. *Biophysical Journal*, 99(8), 2408–2413. https://doi.org/10.1016/j.bpj.2010.08.006

Viswanatha, R., Li, Z., Hu, Y., & Perrimon, N. (2018). Pooled genome-wide CRISPR screening for basal and context-specific fitness gene essentiality in *Drosophila* cells. *Elife*, 7, e36333. https://doi.org/10.1101/274464

Wang, T., Guan, C., Guo, J., Liu, B., Wu, Y., Xie, Z., ... Xing, X.-H. (2018). Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nature Communications*, 9(1), https://doi.org/10.1038/s41467-018-04899-x

Warren, C. L., Kratochvil, N. C. S., Hauschild, K. E., Foister, S., Brezinski, M. L., Dervan, P. B., ... Ansari, A. Z. (2006). Defining the sequence-recognition profile of DNA-binding molecules. *Proceedings of the National Academy of Sciences*, 103(4), 867–872. https://doi.org/10.1073/pnas.0509843102

Weinreich, D. M. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*, 312(5770), 111–114. https://doi.org/10.1126/science.1123539

Whitehead, T. A., Chevalier, A., Song, Y., Dreyfus, C., Fleishman, S. J., De Mattos, C., ... Baker, D. (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nature Biotechnology*, 30(6), 543–548. https://doi.org/10.1038/nbt.2214

Wilke, C. O., & Christoph, A. (2001). Interaction between directional epistasis and average mutational effects. *Proceedings of the Royal Society B: Biological Sciences*, 268(1475), 1469–1474. https://doi.org/10.1098/rspb.2001.1690

Wilke, C. O., Lenski, R. E., & Adami, C. (2003). Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. *BMC Evolutionary Biology*, 14.

Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, 13(1), 59–69. https://doi.org/10.1038/nrg3095

Wloch, D. M., Szafraniec, K., Borts, R. H., & Korona, R. (2001). Direct estimate of the mutation rate and the distribution of fitness effects in the yeast saccharomyces cerevisiae. *Genetics*, 159, 441–452.

Phillips, P. C., Otto, S. P., & Whitlock, M. C. (2000). Beyond the average: The evolutionary importance of gene interactions and variability of epistatic effects. In J. B. Wolf, E. D. Brodie, & M. J. Wade (Eds.) *Epistasis and the Evolutionary Process* (pp. 20–38). Oxford, UK: Oxford University Press.

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, 8(3), 206–216. https://doi.org/10.1038/nrg2063

Wrenbeck, E. E., Azouz, L. R., & Whitehead, T. A. (2017). Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nature Communications*, 8, 15695. https://doi.org/10.1038/ncomms15695

Wylie, C. S., & Shakhnovich, E. I. (2011). A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proceedings of the National Academy of Sciences*, 108(24), 9916–9921. https://doi.org/10.1073/pnas.1017572108

Xu, L., Barker, B., & Gu, Z. (2012). Dynamic epistasis for different alleles of the same gene. *Proceedings of the National Academy of Sciences*, 109(26), 10420–10425. https://doi.org/10.1073/pnas.1121507109

Yamamoto, N., Nakahigashi, K., Nakamichi, T., Yoshino, M., Takai, Y., Touda, Y., … Mori, H. (2009). Update on the Keio collection of Escherichia coli single-gene deletion mutants. *Molecular Systems Biology*, *5*, 335, https://doi.org/10.1038/msb.2009.92

Ye, P., Peyser, B. D., Pan, X., Boeke, J. D., Spencer, F. A., & Bader, J. S. (2005). Gene function prediction from congruent synthetic lethal interactions in yeast. *Molecular Systems Biology*, *1*(1), E1–E12. https://doi.org/10.1038/msb4100034

Zeitoun, R. I., Garst, A. D., Degen, G. D., Pines, G., Mansell, T. J., Glebes, T. Y., … Gill, R. T. (2015). Multiplexed tracking of combinatorial genomic mutations in engineered cell populations. *Nature Biotechnology*, *33*(6), 631–637. https://doi.org/10.1038/nbt.3177

Zeitoun, R. I., Pines, G., Grau, W. C., & Gill, R. T. (2017). Quantitative tracking of combinatorially engineered populations with multiplexed binary assemblies. *ACS Synthetic Biology*, *6*(4), 619–627. https://doi.org/10.1021/acssynbio.6b00376

Zhang, R., & Lin, Y. (2009). DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Research*, *37*(Database), D455–D458. https://doi.org/10.1093/nar/gkn858

Zwart, M. P., Schenk, M. F., Hwang, S., Koopmanschap, B., de Lange, N., van de Pol, L., … de Visser, J. A. G. M. (2018). Unraveling the causes of adaptive benefits of synonymous mutations in TEM-1 β-lactamase. *Heredity*, https://doi.org/10.1038/s41437-018-0104-z