Review

Cell PRESS

Molecular mechanisms of epistasis within and between genes

Ben Lehner

European Molecular Biology Laboratory–Centre for Genomic Regulation (EMBL-CRG) Systems Biology, the Catalan Institute of Research and Advanced Studies (ICREA), Centre for Genomic Regulation and the Pompeu Fabra University (UPF), c / Dr Aiguader 88, Barcelona 08003, Spain

'Disease-causing' mutations do not cause disease in all individuals. One possible important reason for this is that the outcome of a mutation can depend upon other genetic variants in a genome. These epistatic interactions between mutations occur both within and between molecules, and studies in model organisms show that they are extremely prevalent. However, epistatic interactions are still poorly understood at the molecular level, and consequently difficult to predict *de novo*. Here I provide an overview of our current understanding of the molecular mechanisms that can cause epistasis, and areas where more research is needed. A more complete understanding of epistasis will be vital for making accurate predictions about the phenotypes of individuals.

Why do mutations have different outcomes in different individuals?

Many mutations, for example 'disease-causing' mutations, do not have the same effects in all individuals. One reason for this is the influence of the environment: diet, temperature, pathogens and other risk factors can be major influences on the outcome. An additional possible cause of phenotypic variation can be stochastic molecular or epigenetic variation among individuals [1,2]. However, another major influence on mutation outcome is genetics-the genetic background (or 'genetic environment') can determine whether a mutation affects the phenotype of an individual or not. This is not only true across genetically divergent organisms – a mutation that cause disease in one species can have no effect in another [3–5]-but also within a species – a gene can be essential, for example, in one strain of yeast but not in another strain from the same species [6].

This dependence of mutation outcome on genetic background is, in the broadest sense, what is meant by the term 'epistasis' or 'genetic interaction'. The term epistasis was first introduced by Bateson to refer to the masking of one mutation by another [7], but has evolved to refer (simultaneously) to cover a range of related phenomena (Box 1). Several excellent reviews have discussed the alternative definitions of epistasis, how epistatic interactions are detected, their prevalence, how they can be classified, their importance for evolution, and the use of epistasis as a tool for understanding biology [8–11]. By contrast, the aim of this review is to focus on the molecular biology that underlies epistasis. What are the molecular mechanisms that can cause epistatic interactions between mutations? Why, at the molecular level, does a particular mutation have a different outcome in a different individual? Can we predict which mutations will show epistatic interactions and how? Although some progress has been made, our knowledge of the mechanisms that can cause epistasis is still very incomplete, and our ability to predict interactions *de novo* almost non-existent.

Epistatic interactions can be both alleviating (i.e. suppressive, with a better than expected outcome, alternatively referred to as positive or antagonistic epistasis) or aggravating (i.e. increasing severity, with a worse than expected outcome: negative or synergistic epistasis). Interactions can also occur between sequence variants in the same gene ('intramolecular epistasis') and between variants in different genes ('intermolecular epistasis'). For example, in the case of possible compensatory mutations in a bacteriophage about 50% are estimated to be intramolecular and half intermolecular [12]. In this article I first review what is known about the molecular mechanisms that can underlie intermolecular epistasis, before addressing the problem of interactions within the same molecule. Finally I briefly highlight the context-dependence of epistasis – that is, the importance of higher-order interactions.

Epistasis between genes

Individuals of a species typically differ in the sequences of thousands of different proteins, not to mention variation in non-coding regions of the genome [13,14]. As a result, the consequence of a particular mutation has the potential to be modified by variation in the activities of thousands of different proteins. Experiments using natural, engineered and selected mutations have identified numerous examples of epistasis between genes affecting many different phenotypic traits [15–17], including between natural genetic variants in yeast [18–20], plants [21], nematodes [22,23], flies [24,25], birds [26], and mice [27–29], to highlight a few.

For example, a recent study dissected the genetic causes of variation in sporulation efficiency (the developmental program leading to meiosis) between an oak-tree strain and a vineyard strain of the yeast *Saccharomyces cerevisiae* [19]. Nearly all of the genetic variation was accounted for by four nucleotide variants affecting three different transcription factors. Introducing the four alleles individually and in combination into the different strains revealed

Corresponding author: Lehner, B. (ben.lehner@crg.eu).

Box 1. Defining epistasis

The term epistasis has a plurality of meanings [10,11], but was first used by Bateson to describe the masking of the effects of one genetic variant by another [7]. This is still the sense in which many geneticists use the term [62]. Later, R.A. Fisher used the term 'epistacy' to refer to any statistical deviation from the additive combination of two loci [107], but epistasis is now adopted in the population genetics literature to replace this term. When referring to fitness, this is probably the most useful definition for evolutionary theory. The systems biology and genomics communities tend to use the word in a similar sense, but often referring to phenotypes rather than to fitness, and quantifying deviation against a single reference genotype rather than the average deviation across all genotypes in a population [11]. Further, the 'expected' phenotype (for example growth) is often empirically defined using data from the very large numbers of mutation combinations assayed [34,37]. This community also tends to use the term 'genetic interaction' interchangeably with epistasis. For human genetics, where we are interested in the particular outcome of a mutation in a particular individual (i.e. patient), it is this last definition that is perhaps most useful. All of these subtly different meanings of the word are still in wide usage, and this can create problems when translating findings between fields [10,11]. In this article I use 'epistasis' in its broadest sense to refer to the dependence of the outcome of a mutation on the genetic background.

strong epistatic interactions between these polymorphisms; these interactions were found to be important for determining the reduced sporulation efficiency of the vineyard strain [19].

Similar pervasive epistasis was also detected in a study that used thousands of gene-expression levels as quantitative traits [30]. Using data from a cross between two yeast strains it was estimated that the effects of more than half of primary quantitative trait loci (QTLs) affecting gene expression are modified by interactions with secondary loci. Strikingly, most of these secondary loci had individual effects too small to be detected in a genome-wide scan [30]. This suggests that many important epistatic interactions will occur with polymorphisms that might not themselves have a significant statistical association with a disease or phenotype in a genome-wide scan. This is an extremely important lesson that should be taken into account when designing human genetic-association studies.

However, the best evidence for the pervasive nature of epistasis (at least with respect to the strong loss-of-function mutations that have been studied) perhaps derives from large-scale reverse genetic screens [31,32]. Here, pairs of mutations (or RNA interference treatments) are systematically combined, and the effects on viability or growth are then determined [33–38]. The negative and positive interactions between millions of different loss-of-function mutations in yeast have been quantified using this approach [34]. The results of these and similar screens in *Caenorhabditis elegans* [33,35] show that for nearly all strong loss-of-function mutations, the effects of the mutation can be influenced by perturbing the activity of many additional genes [31,32]. It should be noted that, to date, mutations with weaker effects on protein function have been less well investigated in these systematic studies, and it remains to be established whether similar pervasive epistasis will be detected for these. However, the results from quantitative trait studies strongly suggest that this will be the case [15,17].

Many thousands of intermolecular epistatic interactions have been identified in the laboratory, but the molecular mechanisms that cause these interactions are in most cases unknown. Indeed, because multiple molecular mechanisms can underlie similar epistatic interactions, the detection of an interaction alone provides little information about the underlying molecular mechanism [11,39]. Although much more work is needed in this field, I summarize here some of the molecular interactions that could cause epistatic interactions between mutations in two different molecules (Figure 1).

Molecular mechanisms of epistasis between molecules

Interactions between changes in interaction interfaces

Perhaps the simplest molecular mechanism that can cause epistasis between two genes is if their two protein products directly interact. For example, a mutation in one protein that affects a physical interaction can be compensated for by a mutation in the interaction partner. In *Salmonella typhimurium* the detrimental effects of antibiotic-resistance mutations in the S12 protein of the small ribosomal subunit can be compensated for by mutations in the L19 protein of the large subunit. L19 is one of the few large subunit proteins that contacts the small subunit [40]. As a second example, the proteins encoded by two physically interacting sex-determining genes in *Caenorhabditis* species, *fem-3* and *tra-2*, only physically interact with the partner from the same species because of rapid compensatory molecular evolution [41]. Similar changes in molecular recognition could also underlie synergistic epistasis, for example a mutation might only be detrimental following a change in the interaction interface of an interaction partner. More complex scenarios can also be envisaged that involve competition between interaction partners. For example, a mutation that increases the affinity of one interaction might have a detrimental effect on a competing interaction unless the affinity of that interaction also increases.

Functional redundancy

A second simple cause of negative epistasis is functional redundancy. If two or more genes perform a common molecular function, then loss of one gene will normally only have a modest effect on that function. Genetic redundancy can easily arise through gene duplication, and gene duplicates are indeed very likely to have strong negative epistatic interactions [42–44]. However, the first systematic genetic interaction screens quickly revealed that interactions between duplicates only account for a small subset of all negative epistatic interactions [38,45–47].

Even in the cases of epistatic interactions between duplicates, the underlying molecular mechanisms can be complex, with compensation depending upon both partial functional redundancy and regulatory interactions between the duplicates [48]. This can give rise to what has been termed 'mixed epistasis' with different sets of target genes showing distinct patterns of expression change across single and double mutants [49]. Complete or partial redundancy is normally only observed for a subset of target genes, with other targets showing distinct patterns of expression changes [49]. It has been proposed that mixed

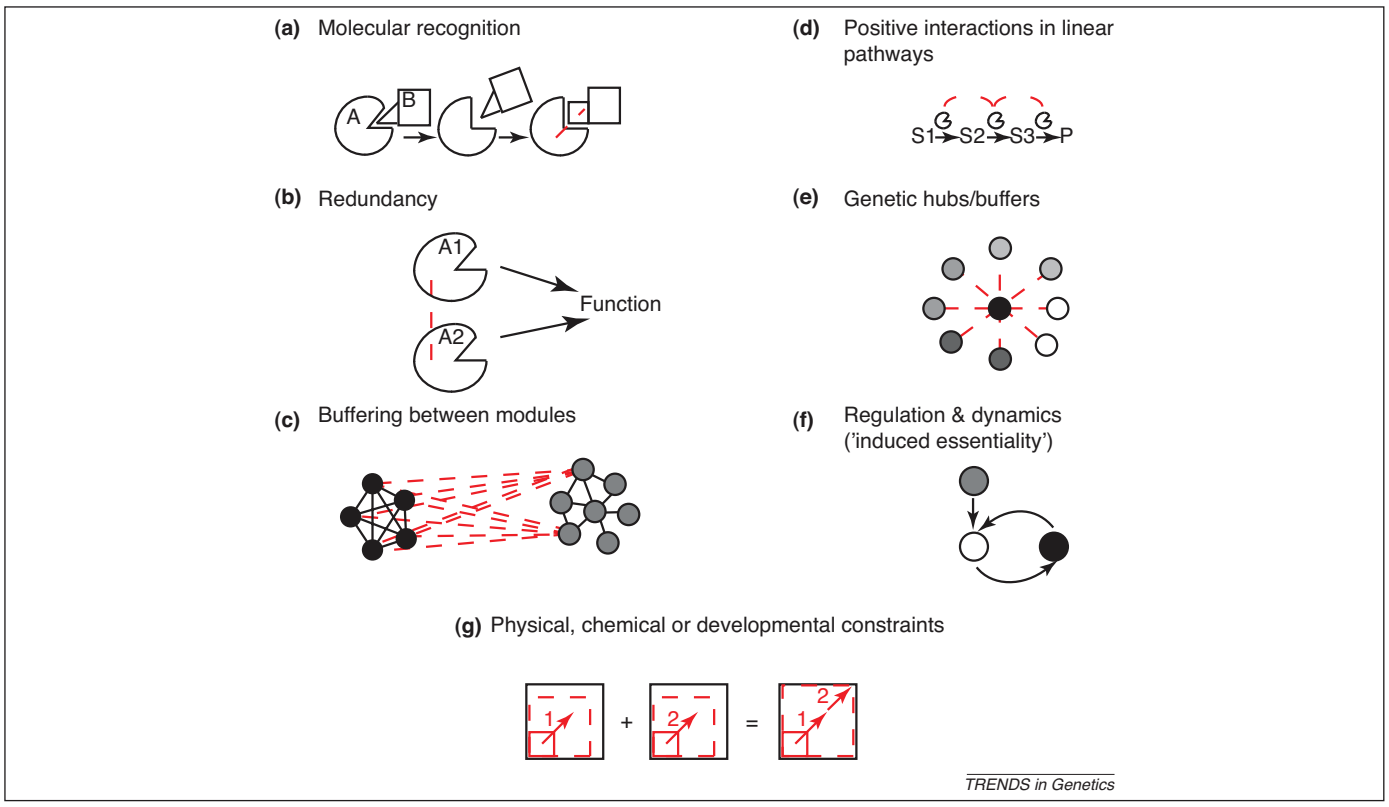

Figure 1. Molecular mechanisms of epistasis between genes. Epistatic interactions (dashed red lines) between mutations in different genes can be caused by diverse underlying molecular mechanisms. **(a)** If two genes encode physically interacting proteins, then a mutation in the interaction interface of one protein (A) could be compensated for by a complementary mutation at the interface of the interaction partner (B). **(b)** If two genes, for example two ancestral duplicates (A1 and A2), perform a common function, then the loss of one gene can be compensated for by continued activity of the second gene. Both genes must be inactivated to lose the function, resulting in strong negative epistasis. **(c)** Many epistatic interactions identified in yeast are enriched between distinct cellular modules (or pathways or complexes), suggesting that these modules buffer each other or work together, although in most cases the precise molecular mechanism is not clear. **(d)** In a linear molecular pathway, such as the metabolic pathway shown here converting a substrate S1 into an essential product P, the inactivation of any gene in the pathway will result in a failure to produce the product. As a result further inactivating mutations in the pathway will have no additional effect, resulting in positive epistasis. **(e)** In genome-scale genetic interaction networks a subset of 'hub' or 'buffer' genes (black nodes) have many epistatic interactions with genes with diverse functions (other nodes). Examples include chaperones that promote the folding or activity of many proteins and chromatin-remodeling complexes that influence the expression of many loci or limit the effects of stochastic and environmental variation. **(f)** Many epistatic interactions could also be caused by non-linear regulatory dynamics. For example, feedback in gene regulatory networks can result in non-trivial outcomes when mutations are combined. Here nodes are genes and arrows represent regulatory interactions. **(g)** Finally, combinations of mutations could have less than the expected outcomes because of physical, chemical or developmental constraints on a system. For example, two mutations (1 and 2) that alone increase the size of an organ (red square) within an animal (black square) here have less than the expected additive effect in combination because of size constraints imposed by the animal.

epistasis reflects the use of duplicates to control sets of downstream genes in both a coupled and an uncoupled manner, depending upon the conditions [49,50].

Interactions between pathways or functional modules

The concept of redundancy can be extended beyond single genes to molecular pathways and modules. For example, if in a particular condition two metabolic pathways can produce the same metabolite, then inactivation of either pathway alone will have little effect on fitness [51]. Consistent with this, genes with similar molecular functions normally have similar profiles of epistatic interactions, suggesting that a seed set of epistatic interactions can be used to accurately predict more interactions for a gene [34,38,52–57].

Two very open questions concern why particular pathways or modules in a cell interact epistatically, and to what extent this reflects (functional) 'redundancy' in the cell. In some cases these interactions might indeed reflect an underlying redundancy (or 'buffering') between pathways, for example between two alternative pathways to degrade an mRNA [53]. However, many interactions are detected between seemingly unrelated cellular processes [38], such as interactions in yeast between the conserved oligomeric Golgi (COG) complex and RNA polymerase II [34,52]. Indeed it has been suggested that many of these 'between-pathway' interactions are actually caused by the regulatory responses of a cell to gene inhibition, rather than by simple redundancy [58]. Understanding how the regulatory responses to a mutation can cause another gene to become essential (the 'induced essentiality' model [58]) poses a major challenge (see below). However, irrespective of the underlying mechanism, the ability to map epistatic interactions onto functional pathways does mean that additional interactions can be readily predicted from a set of seed interactions (Box 2).

Antagonistic interactions in linear pathways

In contrast to negative interactions, it has been proposed that positive epistasis could occur more frequently between genes that function in the same non-essential molecular complex, or in the same pathway or module [34,37,52,59–61]. The basis for this expectation is the following logic derived from metabolism: if an essential metabolite is

Box 2. Predicting epistatic interactions between genes

Most methods for predicting epistasis have focused on analyzing potential synergistic interactions between different genes, or have only predicted the presence or absence of an interaction, and not its direction or strength. Many approaches are based on the principle that functionally-related genes have similar profiles of epistatic interactions [38,53,108]. This means that new interactions can be predicted from a set of 'seed' interactions: genes highly connected by physical or functional interactions with a seed set should also interact epistatically with a gene of interest within that set. Interactions used to make predictions in this way can be physical [55,56], genetic [109], or 'functional' (that is, predicting participation in a common process: functional interactions can be inferred using many different data types [54]). Epistatic interactions are also enriched between functionally-related genes, and functional-relatedness can also be used to predict new interactions [38,110]. Interactions have also been predicted for metabolic enzymes using flux-balance analysis [51,93]. One recent study aimed at the more ambitious goal of predicting epistatic interactions from a gene regulatory network. The study employed matrix decomposition of gene expression changes in single and double mutants and existing molecular interactions from many different datasets to infer a linear gene-regulatory network for the filamentous growth response of yeast [73]. Based on this network it was possible to predict successfully whether 13 new double-mutant combinations resulted in an enhanced or reduced filamentation phenotype. In 6/13 cases the relative phenotypes of the wild-type, single-mutant, and double-mutant strains were also correctly predicted. Further efforts in this direction, including the use of dynamic regulatory models and models that allow feedback, will probably be necessary to predict quantitative epistatic interactions accurately.

produced by a linear metabolic pathway, then the inactivation of any gene in that pathway will prevent the production of that metabolite. This means that (in a strictly linear pathway) the inactivation of a second gene in the pathway can have no further consequence. Thus, in linear pathways null mutations are expected to interact with strong antagonistic (positive) epistasis. Similar masking interactions can occur in signal transduction pathways, for example between mutations with opposing influences on pathway activity. For example the inactivation of an upstream signaling protein can have no effect if a downstream effector is constitutively activated. Indeed this 'classical' (Batesonian) epistasis has been used very successfully to order genes into pathways, often before these pathways were biochemically characterized [62].

However, although the classification of positive and negative epistatic interactions into 'within' and 'between' pathway mechanisms is appealing, in reality most positive genetic interactions still take place between different complexes and modules rather than within the same module [34,52,63]. The molecular mechanisms that cause these interactions are not well understood.

Genetic hubs and buffers

Although most genes make few strong negative epistatic interactions, a subset of genes interact with many different loci, and with loci that have many different functions. Genes with a large number of interactions have been variously termed genetic hubs, buffers or capacitors, and are strikingly biased in their molecular functions. For example, in both *C. elegans* [35] and yeast [34], many of the most connected genes in genetic interaction networks function in the regulation of chromatin. Further, the genome-scale yeast genetic interaction network revealed that genes encoding proteins that function in the secretory pathway also have many negative genetic interactions [34]. Chaperones that promote protein folding and activity, such as Hsp90, can also be classed as genetic hubs, because their inactivation enhances the effects of many different sequence variants across different species [64–66].

In general it is not understood why certain genes behave as genetic hubs. Certainly these genes have the potential to influence the activity of many different proteins, for example in the case of chaperones by assisting protein folding. As an overall trend, there is also a strong correlation between the fitness of a single mutant and the number of epistatic interactions detected for that mutation [34,52,67]. The genes that act as genetic hubs are also normally important for buffering both environmental change and stochastic (i.e. random) variation [34,65,68,69]. The ability of certain genes to act as promiscuous buffers of genetic variation might, therefore, be a byproduct of their importance in ensuring environmental and stochastic resilience, otherwise known as canalization [68,70].

Regulatory dynamics and induced essentiality

It is possible that many epistatic interactions between mutations actually reflect the non-linear dynamics of regulatory networks. Cooperative interactions [71], feedback [39], and complex networks of regulation are frequent in biology and can all lead to non-intuitive responses when mutations are combined [39,72]. To date few studies have attempted to understand how regulatory dynamics can cause epistasis [73,74]; much more work is needed to explore this idea.

Physical constraints

Finally, one possible cause of epistasis (both within and between molecules) that is often overlooked is a physical, chemical or developmental constraint on a system. If a trait has a finite maximum or minimum value, then combinations of mutations can have a smaller effect than the additive expectation, simply because this maximum or minimum effect is reached. Intuitively, many phenotypic traits in biology are likely to have physical bounds. For example, the combination of multiple mutations that individually double the size of an organ is unlikely to result in an unlimited increase in organ size simply because of space constraints.

Epistasis within a gene

Systematic studies using engineered mutations and QTL mapping projects have (for technical reasons) primarily identified epistatic interactions between variants in different genes. Epistasis can, however, also occur between mutations within the same molecule. These intramolecular interactions have been studied more by researchers interested in protein engineering and evolution [4,75–82]. In contrast to intermolecular epistasis, many known examples of intramolecular epistasis therefore involve mutations that combine to produce new functions or increase fitness, rather than decreasing it. As a result, most of these

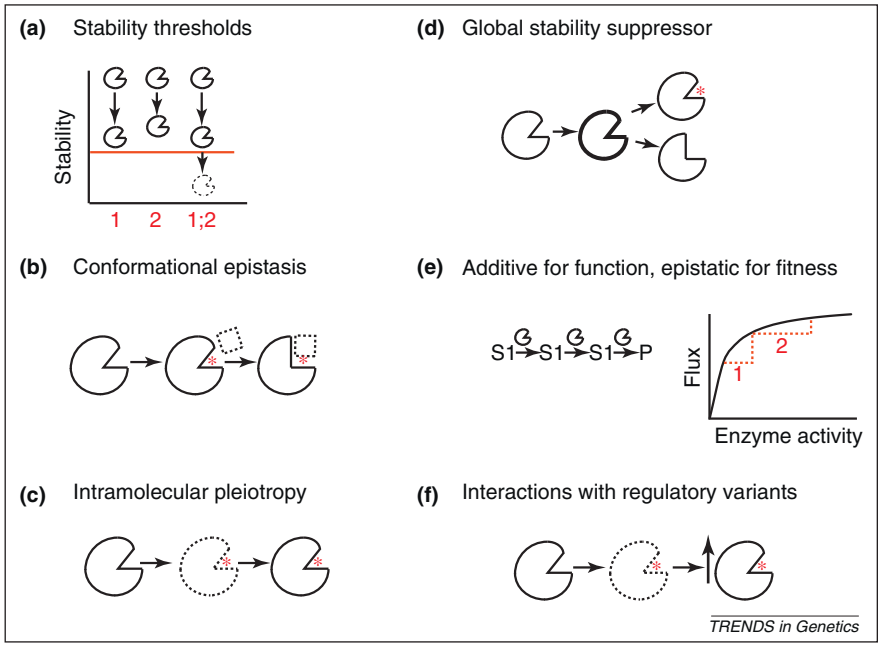

Figure 2. Molecular mechanisms of epistasis within a gene. Epistasis can also occur between mutations in the same gene or locus. **(a)** For example, threshold effects in protein stability can cause negative epistatic interactions. If a protein has 'redundancy' in its stability, meaning that mutations only have a detrimental effect if they reduce stability below a critical threshold (red line), then two mutations alone (1 and 2) could have little effect, but in combination (1;2) be very detrimental. **(b)** Synergistic epistasis can also occur when a conformation change is required for a beneficial (or detrimental) mutation to realize its effect on protein function. Here a beneficial mutation (red asterisk) has no effect until a second mutation that causes a conformation change allows the mutated residue to contact a novel substrate (dashed square). **(c)** Many mutations are likely to have pleiotropic (i.e. multiple) effects on protein activity. For example, a mutation might have a beneficial effect for one function but a detrimental consequence for another, and therefore reduce fitness unless a second compensatory change occurs. Here the example of a potentially beneficial mutation (red asterisk) that reduces protein stability is shown. The mutation also destabilizes the protein and so must be compensated for by a second mutation to realize its beneficial effects on fitness. **(d)** Some mutations could act as global suppressors of detrimental mutations. For example, in the case of protein stability a mutation that increases protein stability (represented by the thick line) could compensate for diverse destabilizing but potentially beneficial mutations (indicated by red asterisk or conformation change). **(e)** Genes do not function in isolation, and mutations causing a linear increase in gene activity will often result in a non-linear increase in fitness. Here two mutations (1 and 2) that double (or halve) the activity of an enzyme do not cause a fourfold increase (or decrease) in metabolic-pathway flux because of a concave activity–flux relationship. **(f)** Finally, epistatic interactions could also occur between mutations that affect the sequence of a protein or RNA and mutations that affect the regulation of that gene. For example, a destabilizing mutation (red asterisk) could be compensated for by a mutation in a regulatory region that increases protein expression (thick black arrow).

interactions are compensatory, whereas most systematically mapped interactions between molecules are synergistic [34].

In both RNAs [83] and proteins [75,79,82] the effects of one mutation can depend on mutations at other positions in the same macromolecule. For example, mutations in an enzyme can have little individual effect on activity, but can have dramatic consequences in combination [75]. As for between-molecule epistasis, these interactions could bias evolutionary paths because particular combinations of mutations have low fitness [82].

As an example, a recent study introduced 168 naturally occurring amino acid substitutions from *Pseudomonas aeruginosa* into the *leuB* gene of *Escherichia coli* [78] and 63 of these changes were found to be detrimental. All of these variants must be compensated for in *P. aeruginosa*, because the structure and function of the enzyme are conserved. Indeed, for the detrimental effects of one tested mutation, four of the other substitutions provided partial compensation [78].

Together with intermolecular interactions, pervasive intramolecular epistasis [12] might explain why even ancestral proteins continue to diverge in sequence: the set of tolerated mutations for any protein changes during evolution even if the function of that protein does not itself change [81]. As for intermolecular epistasis, several molecular mechanisms have been proposed to account for intramolecular interactions. Some of these within-gene mechanisms are conceptually fairly similar to between-gene mechanisms (Figure 2).

Molecular mechanisms of epistasis within a gene

Stability thresholds

Threshold epistasis is a term that has been used to describe the situation when a protein has a critical threshold for stability, and multiple sequence changes are required to cross this threshold [75]. Threshold epistasis is conceptually similar to redundancy: in a particular condition a protein has 'redundant' stability and individual mutations might not exhaust this stability margin. Threshold epistasis has been proposed to explain synergistic interactions between mutations in the bacterial antibiotic-resistance enzyme β-lactamase (TEM-1): the protein has excess or redundant stability, and several mutations are required to have a crucial effect on folding [75]. It is interesting to speculate that, similar to redundancy, threshold epistasis might function to ensure robust protein function in the face of stochastic or environmental change [68].

Conformational epistasis

A second type of intramolecular epistasis can arise in changes in protein conformation. An example could have occurred during the evolution of the glucocorticoid receptor. Here a change in ligand specificity resulted from a pair

of mutations: one of these introduced a residue that ultimately interacted with the new ligand, and the second caused a conformation change that repositioned the first residue such that it is able to contact the ligand [79]. Such changes are highly epistatic – the crucial residue must be both mutated and repositioned to allow binding of the new ligand. It is reasonable to envisage that many potentially beneficial mutations will similarly rely on conformation changes before they can alter the function of a protein.

Intramolecular pleiotropy

Other examples of epistasis occur because mutations can have multiple different (i.e. pleiotropic) effects on a protein. As mentioned above, function-altering mutations can also have the side-effect of reducing protein stability [82,84]. For example, the introduction of a polar residue into an active site can disrupt hydrophobic contacts and so reduce the thermodynamic stability of a protein. Such pleiotropic effects can render a beneficial mutation detrimental in the absence of compensatory changes. This is the case for antibiotic resistant isoforms of TEM-1, where secondary substitutions far from the active site are required to compensate for the destabilizing effects of changes in the active site [85]. Similar trade-offs could also occur beyond effects on stability, for example mutations could be beneficial to one function of a protein but detrimental to another and so require compensatory changes.

Global suppressor mutations

Just as a genetic hub gene can be a promiscuous intermolecular suppressor of genetic variation, so can mutations within a molecule. In the case of protein stability, a mutation that increases stability could act to suppress many different destabilizing but function-altering mutations. In the case of TEM-1, at least ten mutations have been identified that act as 'global' suppressors of destabilizing mutations [75].

Additive for function, epistatic for fitness

It is important to note that combinations of mutations within a gene can additively affect protein function but still interact epistatically to alter fitness or a phenotype [72]. This can occur in metabolism, for example, as a result of nonlinear responses in metabolic flux [86]. *In vivo*, enzymes do not act in isolation, but are kinetically linked via their substrates and products. As a result, changes in catalytic activity at one step in a pathway can be buffered by the response of other enzymes; pathway flux therefore tends to be a concave (hyperbolic) function of enzyme activity [86]. Two successive doublings in enzyme activity can therefore cause a less than fourfold change in pathway flux and fitness [87]. Similar effects might occur in signaling and transcription networks, such that linear changes in the activity of a single protein do not translate to linear increases in fitness.

Interactions with regulatory regions

It is also worth noting that interactions could occur between coding mutations and mutations affecting non-coding regulatory regions. For example, a mutation that reduces activity might have no effect if a second mutation results in increased protein expression [88,89]. This is the case for antibiotic resistance mutations in *Salmonella enterica* isoleucyl-tRNA synthetase that reduce growth rate: both amplification of the gene and mutations that improve the promoter compensate for this fitness cost [89]. In general epistatic interactions with regulatory variants have been much less explored experimentally, but warrant attention.

Higher-order effects: epistatic interactions depend on both the environment and the genetic context

The molecular mechanisms that can cause epistatic interactions are therefore diverse, and still remain to be fully explored in future work. One additional important observation is that epistatic interactions are themselves often highly context-dependent: the particular interaction between two mutations can depend upon both the environment and additional genetic variation.

Environmental influences on epistasis have been observed in phage [90], bacteria [91] and yeast [43,61,92,93]. For example, the interactions between polymorphisms influencing sporulation efficiency in yeast change across environmental conditions [94]. Similar changes in both the magnitude and direction of epistasis have been predicted using metabolic-flux balance analysis [93,95]. Interestingly, the interactions specific to a particular condition could, in general, be more informative about the biology of that condition than the interactions shared across conditions, primarily because of promiscuous interactions with highly connected genes that are conserved across conditions [92].

The extent and direction of epistasis can also depend upon the genetic background, and this is the case for the interactions influencing sporulation in yeast [94]. In effect, therefore, not only are pairwise interactions important for determining phenotypic variation, but so too are higher-level interactions involving multiple different genetic variants and the environment [38,96]. What appear to be gene–gene interactions in a particular experiment can become gene–gene–environment–genetic background interactions when additional experiments are performed. The epistasis detectable in an individual could therefore differ substantially from the epistasis detectable from population level measurements. Again, this could seriously complicate the detection of epistasis in human association studies without reducing its importance for the phenotypes of particular individuals.

Consistent with their context-sensitivity, epistatic interactions also appear to evolve fairly rapidly across species, particularly in the case of negative interactions [58,97,98]. However, particular examples of epistasis, such as interactions between gene duplicates, can be preserved over extensive evolutionary periods [99]. Positive genetic interactions within highly conserved protein complexes or modules can also be conserved across species [97,98] because these complexes are themselves highly conserved.

One of the most striking examples of the importance of higher-order interactions was described in a comparison of gene essentiality between two strains of *S. cerevisiae* yeast. By constructing a systematic gene-deletion library, 44 genes were identified that are essential for the viability of the Sigma1278b isolate of *S. cerevisiae* but not of the

standard S288c strain. Remarkably, genetic analysis revealed that in 13/18 tested cases the differences in essentiality were influenced by at least four different loci in the genome [6]. Thus, rather than resulting from a binary epistatic interaction, most changes in gene essentiality result from complex multigenic epistasis involving at least five different loci. It will be important to establish how often such higher-order interactions account for phenotypic variation between individuals in natural populations.

Concluding remarks

It has been argued that phenotypic variation in a population could, in many cases, be accounted for by purely additive genetic models [100]. However, this is only a theoretical possibility [101], and it contradicts both the demonstrated importance of epistasis in particular human diseases [102–105] and the pervasive epistasis that has been detected in model organisms and highlighted here. It is also somewhat inconsistent with patterns of sequence evolution [79,81,106] and inconsistent with our understanding of molecular biology and the abundance of nonlinear regulatory interactions [86]. Put simply, although they are very challenging to predict and detect in human populations because of a lack of statistical power [8,9], from what is currently understood about genetic architecture and biology, epistatic interactions between mutations are likely to be central to what makes us unique, both in health and disease.

Acknowledgments

Our research is funded by a European Research Council Starting Grant, Grant BFU2008-00365 from the Ministerio de Ciencia e Innovación (MICINN), the ICREA, the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), the European Research Area for Systems Biology (ERASysBio+), the EMBL–CRG Systems Biology Program, and the European Molecular Biology Organization Young Investigator Program. I thank Fyodor Kondrashov, three referees and the editor for useful comments and suggestions, particularly one referee for pointing out the importance of physical constraints.

References

1 Fraga, M.F. *et al.* (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10604–10609
2 Raj, A. and van Oudenaarden, A. (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216–226
3 Azevedo, L. *et al.* (2006) Epistatic interactions: how strong in disease and evolution? *Trends Genet.* 22, 581–585
4 Kondrashov, A.S. *et al.* (2002) Dobzhansky–Muller incompatibilities in protein evolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14878–14883
5 Kulathinal, R.J. *et al.* (2004) Compensated deleterious mutations in insect genomes. *Science* 306, 1553–1554
6 Dowell, R.D. *et al.* (2010) Genotype to phenotype: a complex problem. *Science* 328, 469
7 Bateson, W. (1909) *Mendel's Principles of Heredity*, Cambridge University Press
8 Carlborg, O. and Haley, C.S. (2004) Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* 5, 618–625
9 Cordell, H.J. (2009) Detecting gene–gene interactions that underlie human diseases. *Nat. Rev. Genet.* 10, 392–404
10 Phillips, P.C. (1998) The language of gene interaction. *Genetics* 149, 1167–1171
11 Phillips, P.C. (2008) Epistasis-the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9, 855–867
12 Poon, A. and Chao, L. (2005) The rate of compensatory mutation in the DNA bacteriophage phiX174. *Genetics* 170, 989–999
13 Durbin, R.M. *et al.* (2010) A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073
14 Liti, G. *et al.* (2009) Population genomics of domestic and wild yeasts. *Nature* 458, 337–341
15 Flint, J. and Mackay, T.F. (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* 19, 723–733
16 Hartman, J.L.t. *et al.* (2001) Principles for the buffering of genetic variation. *Science* 291, 1001–1004
17 Mackay, T.F. *et al.* (2009) The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10, 565–577
18 Deutschbauer, A.M. and Davis, R.W. (2005) Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat. Genet.* 37, 1333–1340
19 Gerke, J. *et al.* (2009) Genetic interactions between transcription factors cause natural variation in yeast. *Science* 323, 498–501
20 Heck, J.A. *et al.* (2006) Negative epistasis between natural variants of the *Saccharomyces cerevisiae MLH1* and *PMS1* genes results in a defect in mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3256–3261
21 Kroymann, J. and Mitchell-Olds, T. (2005) Epistasis and balanced polymorphism influencing complex trait variation. *Nature* 435, 95–98
22 Shook, D.R. and Johnson, T.E. (1999) Quantitative trait loci affecting survival and fertility-related traits in *Caenorhabditis elegans* show genotype–environment interactions, pleiotropy and epistasis. *Genetics* 153, 1233–1243
23 Seidel, H.S. *et al.* (2008) Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* 319, 589–594
24 Montooth, K.L. *et al.* (2003) Mapping determinants of variation in energy metabolism, respiration and flight in *Drosophila*. *Genetics* 165, 623–635
25 Gibson, G. *et al.* (2004) Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. *Genetics* 167, 1791–1799
26 Carlborg, O. *et al.* (2006) Epistasis and the release of genetic variation during long-term selection. *Nat. Genet.* 38, 418–420
27 Shao, H. *et al.* (2008) Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19910–19914
28 Stylianou, I.M. *et al.* (2006) Quantitative trait locus analysis for obesity reveals multiple networks of interacting loci. *Mamm. Genome* 17, 22–36
29 Wolf, J.B. *et al.* (2005) Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. *Genetics* 171, 683–694
30 Brem, R.B. *et al.* (2005) Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* 436, 701–703
31 Dixon, S.J. *et al.* (2009) Systematic mapping of genetic interaction networks. *Annu. Rev. Genet.* 43, 601–625
32 Lehner, B. (2007) Modelling genotype–phenotype relationships and human disease with genetic interaction networks. *J. Exp. Biol.* 210, 1559–1566
33 Byrne, A.B. *et al.* (2007) A global analysis of genetic interactions in Caenorhabditis elegans. *J. Biol.* 6, 8
34 Costanzo, M. *et al.* (2010) The genetic landscape of a cell. *Science* 327, 425–431
35 Lehner, B. *et al.* (2006) Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat. Genet.* 38, 896–903
36 Pan, X. *et al.* (2006) A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell* 124, 1069–1081
37 Schuldiner, M. *et al.* (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 123, 507–519
38 Tong, A.H. *et al.* (2004) Global mapping of the yeast genetic interaction network. *Science* 303, 808–813
39 Gjuvsland, A.B. *et al.* (2007) Statistical epistasis is a generic feature of gene regulatory networks. *Genetics* 175, 411–420
40 Maisnier-Patin, S. *et al.* (2007) Compensatory evolution reveals functional interactions between ribosomal proteins S12, L14 and L19. *J. Mol. Biol.* 366, 207–215
41 Haag, E.S. *et al.* (2002) Rapid coevolution of the nematode sex-determining genes *fem-3* and *tra-2*. *Curr. Biol.* 12, 2035–2041

42 Dean, E.J. *et al.* (2008) Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genet.* 4, e1000113
43 Musso, G. *et al.* (2008) The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. *Genome Res.* 18, 1092–1099
44 Tischler, J. *et al.* (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 Biol.* 7, R69
45 Ihmels, J. *et al.* (2007) Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. *Mol. Syst. Biol.* 3, 86
46 VanderSluis, B. *et al.* (2010) Genetic interactions reveal the evolutionary trajectories of duplicate genes. *Mol. Syst. Biol.* 6, 429
47 Wagner, A. (2005) Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays* 27, 176–188
48 Kafri, R. *et al.* (2005) Transcription control reprogramming in genetic backup circuits. *Nat. Genet.* 37, 295–299
49 van Wageningen, S. *et al.* (2010) Functional overlap and regulatory links shape genetic interactions between signaling pathways. *Cell* 143, 991–1004
50 Kafri, R. *et al.* (2009) Genetic redundancy: new tricks for old genes. *Cell* 136, 389–392
51 Papp, B. *et al.* (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429, 661–664
52 Baryshnikova, A. *et al.* (2010) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* 7, 1017–1024
53 Kelley, R. and Ideker, T. (2005) Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.* 23, 561–566
54 Lee, I. *et al.* (2010) Predicting genetic modifier loci using functional gene networks. *Genome Res.* 20, 1143–1153
55 Le Meur, N. and Gentleman, R. (2008) Modeling synthetic lethality. *Genome Biol.* 9, R135
56 Wong, S.L. *et al.* (2004) Combining biological networks to predict genetic interactions. *Proc. Natl. Acad. Sci. U.S.A.* 101, 15682–15687
57 Ulitsky, I. and Shamir, R. (2007) Pathway redundancy and protein essentiality revealed in the *Saccharomyces cerevisiae* interaction networks. *Mol. Syst. Biol.* 3, 104
58 Tischler, J. *et al.* (2008) Evolutionary plasticity of genetic interaction networks. *Nat. Genet.* 40, 390–391
59 Bandyopadhyay, S. *et al.* (2008) Functional maps of protein complexes from quantitative genetic interaction data. *PLoS Comput. Biol.* 4, e1000065
60 Collins, S.R. *et al.* (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446, 806–810
61 St Onge, R.P. *et al.* (2007) Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* 39, 199–206
62 Huang, L.S. and Sternberg, P.W. (2006) Genetic dissection of developmental pathways. *WormBook* 1–19
63 Segre, D. *et al.* (2005) Modular epistasis in yeast metabolism. *Nat. Genet.* 37, 77–83
64 Rutherford, S.L. and Lindquist, S. (1998) Hsp90 as a capacitor for morphological evolution. *Nature* 396, 336–342
65 Queitsch, C. *et al.* (2002) Hsp90 as a capacitor of phenotypic variation. *Nature* 417, 618–624
66 Zhao, R. *et al.* (2005) Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* 120, 715–727
67 Davierwala, A.P. *et al.* (2005) The synthetic genetic interaction spectrum of essential genes. *Nat. Genet.* 37, 1147–1152
68 Lehner, B. (2010) Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. *PLoS ONE* 5, e9035
69 Levy, S.F. and Siegal, M.L. (2008) Network hubs buffer environmental variation in Saccharomyces cerevisiae. *PLoS Biol.* 6, e264
70 Waddington, C.H. (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565
71 Gibson, G. (1996) Epistasis and pleiotropy as natural properties of transcriptional regulation. *Theor. Popul. Biol.* 49, 58–89
72 Wright, S. (1934) Physiological and evolutionary theories of dominance. *Am. Nat.* 68, 24–53
73 Carter, G.W. *et al.* (2007) Prediction of phenotype and gene expression for combinations of mutations. *Mol. Syst. Biol.* 3, 96
74 Gertz, J. *et al.* (2010) Epistasis in a quantitative trait captured by a molecular model of transcription factor interactions. *Theor. Popul. Biol.* 77, 1–5
75 Bershtein, S. *et al.* (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444, 929–932
76 Hinkley, T. *et al.* (2011) A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nat. Genet.* 43, 487–489
77 Kryazhimskiy, S. *et al.* (2011) Prevalence of epistasis in the evolution of influenza a surface proteins. *PLoS Genet.* 7, e1001301
78 Lunzer, M. *et al.* (2010) Pervasive cryptic epistasis in molecular evolution. *PLoS Genet.* 6, e1001162
79 Ortlund, E.A. *et al.* (2007) Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317, 1544–1548
80 Plotnikova, O.V. *et al.* (2007) Conversion and compensatory evolution of the gamma-crystallin genes and identification of a cataractogenic mutation that reverses the sequence of the human *CRYGD* gene to an ancestral state. *Am. J. Hum. Genet.* 81, 32–43
81 Povolotskaya, I.S. and Kondrashov, F.A. (2010) Sequence space and the ongoing expansion of the protein universe. *Nature* 465, 922–926
82 Weinreich, D.M. *et al.* (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312, 111–114
83 Kern, A.D. and Kondrashov, F.A. (2004) Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. *Nat. Genet.* 36, 1207–1212
84 Tokuriki, N. *et al.* (2008) How protein stability and new functions trade off. *PLoS Comput. Biol.* 4, e1000002
85 Wang, X. *et al.* (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* 320, 85–95
86 Kacser, H. and Burns, J.A. (1981) The molecular basis of dominance. *Genetics* 97, 639–666
87 Lunzer, M. *et al.* (2005) The biochemical architecture of an ancient adaptive landscape. *Science* 310, 499–501
88 McLoughlin, S.Y. and Copley, S.D. (2008) A compromise required by gene sharing enables survival: Implications for evolution of new enzyme activities. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13497–13502
89 Paulander, W. *et al.* (2010) Amplification of the gene for isoleucyl-tRNA synthetase facilitates adaptation to the fitness cost of mupirocin resistance in *Salmonella enterica*. *Genetics* 185, 305–312
90 You, L. and Yin, J. (2002) Dependence of epistasis on environment and mutation severity as revealed by *in silico* mutagenesis of phage t7. *Genetics* 160, 1273–1281
91 Remold, S.K. and Lenski, R.E. (2004) Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. *Nat. Genet.* 36, 423–426
92 Bandyopadhyay, S. *et al.* (2010) Rewiring of genetic networks in response to DNA damage. *Science* 330, 1385–1389
93 Harrison, R. *et al.* (2007) Plasticity of genetic interactions in metabolic networks of yeast. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2307–2312
94 Gerke, J. *et al.* (2010) Gene–environment interactions at nucleotide resolution. *PLoS Genet.* 6
95 Snitkin, E.S. and Segre, D. (2011) Epistatic interaction maps relative to multiple metabolic phenotypes. *PLoS Genet.* 7, e1001294
96 van Swinderen, B. and Greenspan, R.J. (2005) Flexibility in a gene network affecting a simple behavior in *Drosophila melanogaster*. *Genetics* 169, 2151–2163
97 Dixon, S.J. *et al.* (2008) Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16653–16658
98 Roguev, A. *et al.* (2008) Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* 322, 405–410
99 Vavouri, T. *et al.* (2008) Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends Genet.* 24, 485–488
100 Hill, W.G. *et al.* (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* 4, e1000008
101 Cheverud, J.M. and Routman, E.J. (1995) Epistasis and its contribution to genetic variance components. *Genetics* 139, 1455–1461

102 de Cid, R. *et al.* (2009) Deletion of the late cornified envelope *LCE3B* and *LCE3C* genes as a susceptibility factor for psoriasis. *Nat. Genet.* 41, 211–215
103 Williams, T.N. *et al.* (2005) Negative epistasis between the malaria-protective effects of alpha$^+$-thalassemia and the sickle cell trait. *Nat. Genet.* 37, 1253–1257
104 Carrasquillo, M.M. *et al.* (2002) Genome-wide association study and mouse model identify interaction between RET and EDNRB pathways in Hirschsprung disease. *Nat. Genet.* 32, 237–244
105 Martin, M.P. *et al.* (2002) Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat. Genet.* 31, 429–434
106 Meer, M.V. *et al.* (2010) Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature* 464, 279–282
107 Fisher, R. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52, 399–433
108 Ye, P. *et al.* (2005) Gene function prediction from congruent synthetic lethal interactions in yeast. *Mol. Syst. Biol.* 1, 0026
109 Qi, Y. *et al.* (2008) Finding friends and enemies in an enemies-only network: a graph diffusion kernel for predicting novel genetic interactions and co-complex membership from yeast genetic interactions. *Genome Res.* 18, 1991–2004
110 Zhong, W. and Sternberg, P.W. (2006) Genome-wide prediction of *C. elegans* genetic interactions. *Science* 311, 1481–1484