Genotype to phenotype: lessons from model organisms for human genetics

Ben Lehner

Abstract | To what extent can variation in phenotypic traits such as disease risk be accurately predicted in individuals? In this Review, I highlight recent studies in model organisms that are relevant both to the challenge of accurately predicting phenotypic variation from individual genome sequences ('whole-genome reverse genetics') and for understanding why, in many cases, this may be impossible. These studies argue that only by combining genetic knowledge with *in vivo* measurements of biological states will it be possible to make accurate genetic predictions for individual humans.

Thousands of genetic variants have now been associated with common human diseases[1,2]. These associations between genetic variation and disease risk have the potential to revolutionize our understanding of common diseases because they identify pathways and processes that are causally implicated in a disease, providing a first step towards the development of targeted therapies and prevention strategies.

However, will a particular individual, carrying a defined set of genetic variants, actually develop one or more of thousands of different diseases? Although often presented as a cornerstone of 'personalized and predictive medicine', making accurate predictions for most common diseases is still an ambitious challenge. Crucially, these predictions must be made at the level of individuals. A patient does not want to know the typical outcome of a mutation that they carry: they want to know what will actually happen to them.

In most cases, our understanding of the genetics of common human diseases is far from complete[2]. Moreover, even with a complete understanding of the genetics of a complex disease, we may never be able to make accurate predictions about disease risk in individuals using genetics alone, as is well demonstrated by the high levels of discordance for most common diseases in identical twins[3–5].

How can we progress to a more complete understanding of the genetics of a disease? And why do even genetically identical individuals often substantially differ in phenotypic traits such as disease risk? The aim of this Review is to highlight recent work in model organisms that is relevant to both of these questions. The goal is not to provide an exhaustive overview but rather to highlight examples of studies that are enriching our understanding of the interplay between genotype and phenotype and so are providing a framework for the development of personalized genetics in humans. I focus in particular on non-vertebrate models, in which much larger-scale systematic experiments have been possible.

I first consider the problem of associating genes and genetic variation with particular phenotypes on a genomic scale. I then proceed to the question of how mutations in multiple genes combine to alter phenotypic traits before introducing the idea of 'whole-genome reverse genetics' to assess our ability to make accurate phenotypic predictions. I then turn to the question of why genome sequences are often insufficient to predict trait variation in individuals. This requires consideration of how the environment, stochastic processes, life history and transgenerational influences interact to determine phenotypic traits in individuals. All of these potentially important influences on phenotypic variation are now being studied at the molecular level in model organisms.

Globally linking genes to phenotypes

Despite the recent explosion of genome-wide association studies (GWASs) in humans, we probably still do not know most of the genetic variants that can influence susceptibility to common diseases[2]. Given a subset of the genes relevant for a trait, how can we predict the rest? This is a question that has been quite extensively investigated in invertebrate models, where systematic forward- and reverse-genetic screens have provided much more complete maps of which genes, when mutated, can influence which phenotypic traits.

Lessons from systematic genetics. In contrast to the situation in mammals, in invertebrate model organisms it is relatively straightforward to carry out systematic genetic

EMBL–CRG Systems Biology, Centre for Genomic Regulation (CRG) and UPF, Carrer Dr. Aiguader 88, 08003 Barcelona, Spain; and Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.
e-mail: ben.lehner@crg.eu
doi:10.1038/nrg3404
Published online
29 January 2013

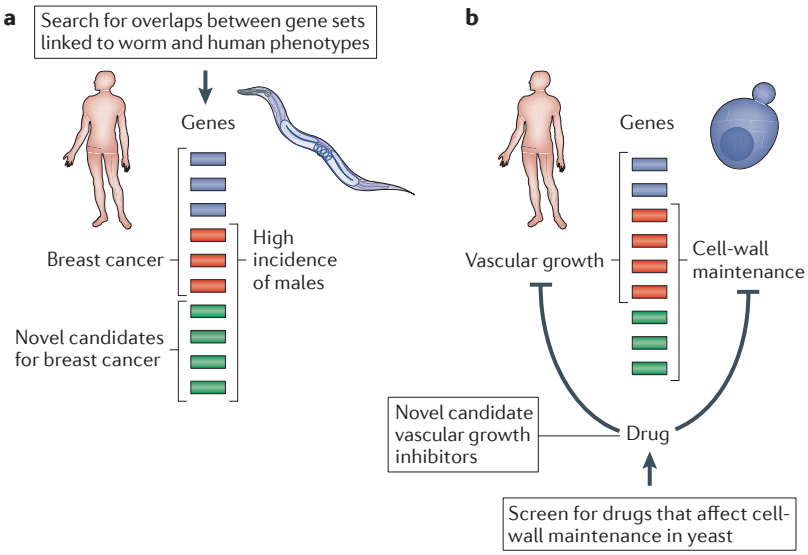

Figure 1 | **Phenologues: mapping phenotypes between organisms.** Perturbation of overlapping modules of orthologous genes may result in one set of phenotypes in one organism but a different set of phenotypes in another organism. **a** | For example, mutations in a module of DNA damage response genes cause breast cancer in humans but a high incidence of males (Him) phenotype in *Caenorhabditis elegans* (owing to chromosome non-dysjunction). Other genes linked to the Him phenotype therefore make good candidate genes for novel breast cancer loci. **b** | The overlap in the sets of genes linked to vascular growth in vertebrates and to cell-wall maintenance in yeast allowed the prediction that the approved antifungal drug thiabendazole would act as an angiogenesis inhibitor[21].

screens. For example, in budding yeast[6], fission yeast[7] and *Escherichia coli*[8], the construction of gene deletion collections means that researchers can 'walk through the genome' identifying all of the gene deletions that influence a phenotype of interest. Similarly, in worms[9] and flies[10], genome-wide RNA interference (RNAi) screens allow the comprehensive identification of genes that influence any trait of interest. Moreover, cheap whole-genome sequencing and genotyping are revolutionizing the ease with which both random laboratory-induced mutations[11] and natural genetic variants[12–16] can be linked to trait variation.

The connections between genes and phenotypes are therefore both more complete and more systematic in model organisms than they are in humans. This provides an unbiased assessment of the genetic complexity of phenotypic traits, and indeed rather than a handful of genes influencing a trait of interest, it is more common to identify hundreds or thousands of genes[9,17–19]. Moreover, genetic screens in model organisms have highlighted that pleiotropy is extremely common: many genes are linked to a wide diversity of traits[9,17–19]. Of course, natural genetic variation in genes may not necessarily be so pleiotropic in consequence but, in general, the pithy statement by Sewall Wright[20] in the 1930s that "each character is affected by many genes and each gene affects many characters" has largely been confirmed by twenty-first century genetics. This seems to be the rule rather than the exception.

Modules
Groups of genes or proteins in a network that have strong interactions among themselves and that carry out particular functions largely independently of other genes or proteins. Mutations in genes from a module often have similar phenotypic consequences.

Orthologous
A gene in one species is orthologous to a gene in another species if they are derived from a common ancestor.

In addition to providing basic insights into genetic architecture, these comprehensive model organism genotype–phenotype maps can also have direct relevance to human disease. This is because genes tend to work in evolutionarily conserved pathways or modules, and so genotype–phenotype maps can be directly transferred between species. For example, mutations in a subset of genes that function in the response to DNA damage tend to cause a high incidence of males (Him) phenotype in *Caenorhabditis elegans* and breast cancer in humans, meaning that new worm *him* genes make good candidates for breast cancer genes[21] (FIG. 1a). These non-obvious relationships between phenotypes in different organisms that are affected by mutations in overlapping sets of orthologous genes are referred to as 'phenologues' and can be systematically identified[21]. Mapping phenologues between species can also predict new clinically relevant drugs. For example, on the basis of the observation that a common set of genes influence cell-wall maintenance in yeast and vascular growth in vertebrates, an approved antifungal drug called thiabendazole was predicted and validated as a novel inhibitor of angiogenesis[22] (FIG. 1b).

Genome-scale networks that link genes to phenotypes. The comprehensive genotype–phenotype data available for model organisms also provide a fantastic resource for developing and evaluating computational methods to predict the connections between genes and phenotypes on a genomic scale.

A powerful strategy to achieve this is 'guilt-by-association': if two genes function in the same pathway or process, then mutations in these genes are likely to have overlapping phenotypic consequences[23,24]. Guilt-by-association is a successful framework because many different types of evidence can be used to identify functionally associated genes (FIG. 2). For example, genes encoding proteins that physically interact, that are co-regulated or that are co-evolving are all more likely to function in a common process. One approach for predicting functionally coupled genes is therefore to integrate evidence from diverse data sets to build large networks of functional associations between genes[25–27]. Building these networks requires the reliability of different data sets to be evaluated and for interactions in the final network to represent a weighted integration of interactions that is inferred from different types of evidence[23,24]. Crucially, the systematic genetic data that are available in model organisms then allow the systematic evaluation of the utility of the networks in guilt-by-association predictions for diverse traits[28,29].

As an example, diverse data sets were used from multiple organisms to construct a network consisting of 102,803 linkages among 5,483 budding yeast proteins (more than 90% of the proteome in this species)[30]. This network was shown to have a broad utility for predicting new genes associated with diverse phenotypic traits[28,31]. Indeed, networks for yeast[25,26,30,32], worms[29,33], mice[34] and plants[35,36] have all been shown to associate thousands of genes accurately to phenotypes. Moreover, similar approaches in humans have shown potential for

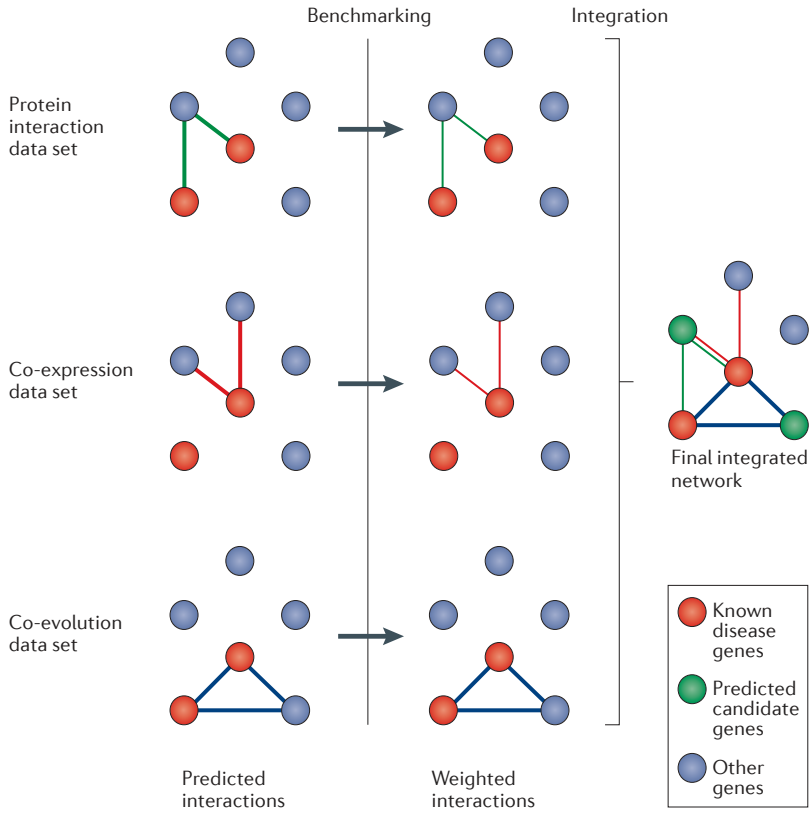

Figure 2 | **Guilt-by-association: integrating data into genome-scale networks that can be used to link genes to phenotypes.** Many different experimental and computational data sets can be used to predict whether two proteins (nodes in the figure) physically or functionally interact (edges in the figure). Predicted functional associations derived from different data sets, such as protein–protein interactions (green lines), co-expression (red lines) and co-evolution (blue lines), can be integrated by first benchmarking the interactions inferred from each data set against a set of 'gold-standard' interactions and then combining them in such a way that the interactions are weighted according to their estimated reliability (weights here are represented by the width of each line). In the final network, proteins are connected by weighted interactions that may derive from one or multiple sources of evidence. Mutations in proteins with high-confidence interactions with known disease genes (red) are predicted candidate genes for the same disease (green).

identifying new genes for Mendelian[27,37] and complex[38] diseases, although much more research is still needed in the application of these ideas to human genetics, especially to GWAS data. Although the guilt-by-association approach may have some bias towards identifying pleiotropic genes[39], extensive experimental validation has demonstrated the practical utility of the approach.

Gain-of-function mutations. To date, large-scale assessments of gene function have largely focused on loss-of-function mutations. However, some systematic data for gain-of-function genetic perturbations are available. For example, in budding yeast, ~15% of protein-coding genes were found to affect growth severely when strongly overexpressed[40]. The properties associated with these 'dosage-sensitive' proteins include: a high content of disordered regions containing linear motifs that are important for protein–protein interactions; a large number of transient protein interactions; and enrichment for protein domains that bind to linear motifs. These features suggest the hypothesis that one of the causes of dosage sensitivity is mass-action-driven promiscuous molecular interactions[41]. However, it is likely that promiscuous ('off-target') molecular interactions are only one cause of dosage sensitivity. Other causes, such as increased or constitutive activation of 'on-target' pathways, protein aggregation, interference in the assembly of protein complexes and disruption of the 'balance' between pathways, have also been suggested to be important[40,42]. The development of techniques that allow the gene copy number at which a gene becomes detrimental to be determined[43] should allow a finer investigation of dosage sensitivity, and in general more screens are needed to link gain-of-function mutations to particular phenotypic traits.

Non-coding variation. An important challenge that also deserves more attention is to predict the effects of mutations in non-protein-coding regions of the genome: for example, when do mutations in regulatory regions affect phenotypic variation? This is particularly important, given the accumulating evidence that many causal variants that influence human disease actually lie outside coding regions and alter gene expression. Systematic maps of transcription factor binding sites, chromatin modifications, chromatin accessibility and expression quantitative trait loci (eQTLs) can be used to pinpoint potentially important regulatory regions[44,45], but this does not directly address the question of whether variation in these regions has phenotypic consequences. To date, evolutionary conservation has been used to identify properties that predict when genetic variation in regulatory regions is most likely to be detrimental[46], but as for gain-of-function protein traits, systematic experimental data on the phenotypic consequences of mutations in regulatory regions is largely missing. The use of large libraries of synthetic promoters[47,48] may provide one starting point, but ultimately comprehensive data sets of links between genetic variation and phenotypic variation will be required to develop and to evaluate predictive methods for regulatory regions.

Systematic analysis of epistasis

Mutations often have consequences that vary across individuals, and one reason for this is epistasis or genetic interactions, which are understood most broadly as the dependence of mutation outcome on genetic background[49,50]. There are numerous examples of epistatic interactions in human disease, and indeed epistatic interactions might, in part, have led to overestimations of the heritability (phenotypic variance attributable to genetic variation) of human disease[51]. Model organisms have been used extensively both to understand epistasis better and to learn how to predict it[52]. Numerous types of epistasis can be envisaged[53], but two important outcomes are that the combined effect of two mutations can be either stronger (negative or synergistic interaction) or weaker (positive or antagonistic interaction) than expected[54].

Disordered regions
Regions of proteins that are intrinsically unfolded; that is, they are without a well-defined tertiary structure under physiological conditions.

Expression quantitative trait loci
(eQTLs). Regions of the genome containing genetic polymorphisms that alter how genes are regulated, influencing how much RNA or protein they produce.

Major- and minor-effect loci
Regions of the genome containing genetic polymorphisms that account for a large or small proportion of variance in a particular phenotype, respectively.

Systematic mapping of genetic interaction networks. Epistatic interactions between mutations have been mapped on a massive scale in budding yeast[55] and on a smaller scale in fission yeast[56,57], *C. elegans*[58,59] and in *Drosophila melanogaster* cells[60]. In yeast, this has been facilitated by the development of selection procedures that allow arrays of mutant strains to be systematically mated to construct double mutants[61]. Genetic interactions have also been studied between natural variants: for example, between QTLs that influence mating efficiency[62,63] and between QTLs that influence gene expression traits (namely, eQTLs)[64]. Here I highlight some important take-home messages from these genetic interaction analyses in model organisms: epistatic interactions are prevalent; genes differ widely in the number of interactions in which they participate; interactions are context-dependent; and interactions can be predicted. These properties seem to be conserved between species, and so they are also likely to apply to human genetic disease.

Epistasis is prevalent. The first important conclusion from these studies is that the potential for genetic interactions is huge: in yeast, significant interactions were detected between ~170,000 different pairs of genes from more than 5 million pairs tested[55]. Put simply, there are many more ways to generate similar phenotypic effects in yeast cells by combining two gene deletions than there are by deleting a single gene. Consistent with this, in an analysis of QTLs influencing gene expression, it was estimated that approximately two-thirds of 225 gene expression traits influenced by two different loci involved a significant interaction between the two loci[65]. Moreover, epistasis between both major- and minor-effect loci was found to be important in determining sporulation efficiency in the progeny of a cross between two budding yeast strains[62,63], and higher-order epistatic interactions involving multiple loci underlie differences in the effects of gene deletions between two laboratory strains of yeast[66].

Disease specifiers and disease modifiers. A second important conclusion from systematic studies of epistasis is that the number of potential genetic interaction partners differs widely among genes: mutations in some genes have many modifier loci but most have far fewer[55]. Genes with many potential genetic interaction partners ('genetic hubs') are functionally biased: for example, they often encode components of chromatin and transcription complexes[55,58]. Genes with many interactions in one species also tend to have many interactions in other species. Moreover, the biased functional properties of hub genes means that they can probably be predicted to some extent even if no orthologous genes exist in a model organism[67]. The same genes and processes that have many genetic interactions in worms and yeast are therefore likely to have many genetic interactions in human disease. One way to consider genetic hubs is as 'disease modifiers'[58]; a mutation in a hub gene has the potential to enhance the effects of mutations in many other loci that alter very different phenotypic traits (FIG. 3).

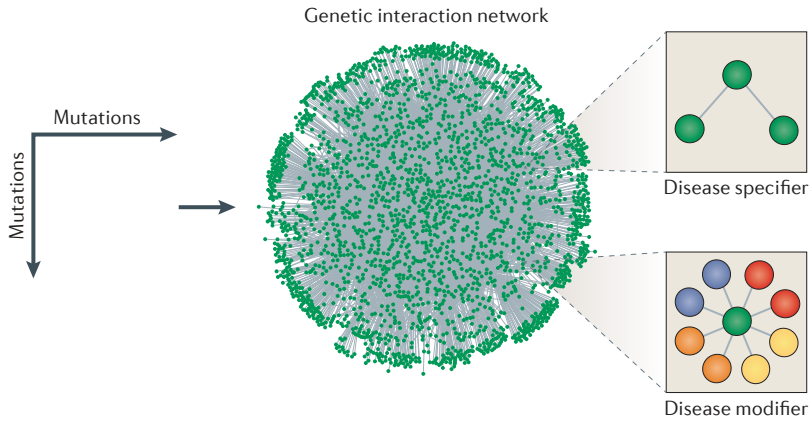

Figure 3 | **Systematic analysis of genetic interactions (epistasis): disease specifiers and disease modifiers.** In yeast, worms and fly cells, the effects of inhibiting two genes simultaneously have been systematically tested for many combinations of genes. This allows large-scale networks to be constructed where each edge in the network represents either a negative (enhanced phenotype) or positive (relieved phenotype) epistatic interaction. Whereas some genes in genetic interaction networks have few genetic interactions with genes of related functions, others interact promiscuously with genes with diverse molecular functions. The first class of genes can be considered to be 'disease specifiers' because their perturbation (alone and in combination) is only likely to influence a limited number of phenotypic traits. By contrast, perturbation of genes with functionally diverse interaction partners (here indicated by differently coloured nodes) may enhance the consequences of mutations in many different processes, depending on the other mutations carried in a genome. These genetic interaction hubs have therefore been termed 'disease modifiers' to reflect this potential for promiscuity[58].

Predicting genetic interactions. The third conclusion from the systematic analysis of genetic interactions is that genes with closely related functions tend to have similar profiles of genetic interactions. This means that genetic interaction profiles can successfully predict gene function[55], but also conversely that gene functions (and interactions that link functionally related genes) can successfully predict genetic interactions[68,69]. In particular, if two genes from two different processes have a negative genetic interaction, this predicts that other genes in the two pathways will also negatively interact[68–72]. Thus, the genetic interaction network is highly modular, and modules of genes share similar profiles of genetic interactions[72]. This allows genetic interactions to be predicted on a genomic scale using guilt-by-association[69,73]. This is important because in higher organisms the systematic experimental mapping of genetic interactions may never be realistic, and the number of possible combinations to test statistically in association studies is immense. Moreover, even in yeast, interactions have primarily been mapped only for a single trait: growth. Thus, extensive efforts are required to learn how best computationally to predict likely interactions that can then be statistically evaluated in human populations.

Genome–environment interactions

In addition to the effects of genetic background, another widely appreciated influence on the outcome of mutations is the environment: mutations may predispose

to a particular disease but only if individuals are also exposed to a particular biotic or abiotic environmental condition or a trigger such as diet, temperature, parental nurturing, variation in the microbiota or exposure to a pathogen.

How mutations interact is also context-dependent. Gene–environment interactions have been studied both systematically and at base-pair resolution in model organisms and have been found to be widespread[17–19,74]. Moreover, epistatic interactions between genes have also been found to be context-dependent. Thus, an interaction detected in one particular condition[74–77] or species[78–80] is often not detected in a second condition or species. This plasticity of genetic interactions predicts that even if a gene is implicated in two particular diseases, it may have different interactions (or modifier loci) in the two pathologies. As such, although the same functional module may be implicated in two diseases, the interactions of this module may differ between one tissue and another.

Promiscuous influences of the environment. In addition to specific interactions with particular mutations, it is important to note that the environment can also influence the outcome of mutations in more general, or promiscuous, ways. For example, the environment can influence the effects of mutations by altering the availability or activity of molecular chaperones, which are proteins that influence the folding or activity of other proteins in the cell[81]. This is because the effects of many mutations are modified by molecular chaperone activity, either because chaperones directly stabilize mutated proteins or because the outcome of a mutation is influenced by the activity of a second, chaperone-dependent pathway or process[82].

The chaperone activity — and therefore the mutation-buffering capacity — of a cell or organism can increase or decrease in response to environmental stimuli. For example, severe environmental stress can titrate away molecular chaperones and can therefore 'unbuffer' (that is, enhance) the effects of otherwise phenotypically inconsequential mutations[83,84]. Conversely, a mild heat shock that induces a protective stress response that includes the induction of chaperones can increase the capacity of an organism to buffer the effects of inherited detrimental mutations[85]. This shows how both current and prior environmental conditions can have promiscuous influences on the outcome of mutations.

Although promiscuous influences of the environment on mutation outcome have largely been considered from the perspective of protein folding, other potential mechanisms may exist by, for example, promiscuous effects on gene expression. These warrant future investigation.

Whole-genome reverse genetics

Ultimately, the best test of our understanding of genetics is whether we can predict phenotypic variation from sequence. In a given environment and for a given phenotype, which individuals will differ and how? This challenge of predicting phenotypic variation from the complete genome sequences of individuals can be termed 'whole-genome reverse genetics'[86]. The goal is to make accurate predictions, preferably for many different phenotypes or conditions simultaneously. This can seem like a daunting task because of the sheer number of genetic variants in each individual. For example, budding yeast strains isolated in different environments or from different regions typically have protein-coding variation in thousands of their 6,000 genes[87]. Simultaneously changing thousands of parameters is might be considered one of the 'worst possible experiments' for a biologist. However, it is a challenge that must be tackled.

Model organisms represent an ideal opportunity for testing our ability to make whole-genome reverse-genetic predictions. In several model organisms, we have fairly comprehensive information from forward- and reverse-genetic screens on the genes that can influence many different phenotypic traits (see above). Moreover, cheap and quantitative experiments can be carried out in model organisms to evaluate prediction performance.

The challenge of making whole-genome reverse-genetic predictions has been attempted and experimentally evaluated in budding yeast[86], but it could also be applied to other model species. In the budding yeast study, only protein-coding variation was considered, and all predictions were made relative to a reference laboratory strain. The approach consisted of three main steps (FIG. 4). First, the variants affecting the amino acid sequence of each protein in each individual were evaluated to determine whether they were likely to alter the function of that protein. There are many different ways to estimate whether particular mutations alter protein function[88], and in this study a fairly simple approach of considering the evolutionary conservation of each amino acid across different yeast species was used. Indeed, on the basis of an evaluation of budding yeast mutations that are known to alter protein function isolated in forward-genetic screens, this evolutionary method seems to work well[86]. Second, for each different environmental condition, the individual yeast strains were ranked using the total function-altering mutation load that they carry in sets of genes that were previously reported to influence growth when deleted under the same condition. These lists of genes were derived from genome-wide screens using the deletion collection. Thus variants in any of these genes were considered to have a similar potential to influence the trait, and their effects were assumed to combine additively. Third, the actual phenotypic differences among individuals were experimentally quantified (in this case, growth rate and efficiency in different environments and resistance to drugs), and these experimental data were used to evaluate the performance of the predictions.

Surprisingly, this simple, protein-only, 'black box' and additive genetic model provided reasonable predictions of phenotypic variation across individuals[86]. Moreover, when predictions failed, this could partially be accounted for by the low reliability of the sets of genes that were reported as influencing the trait, as

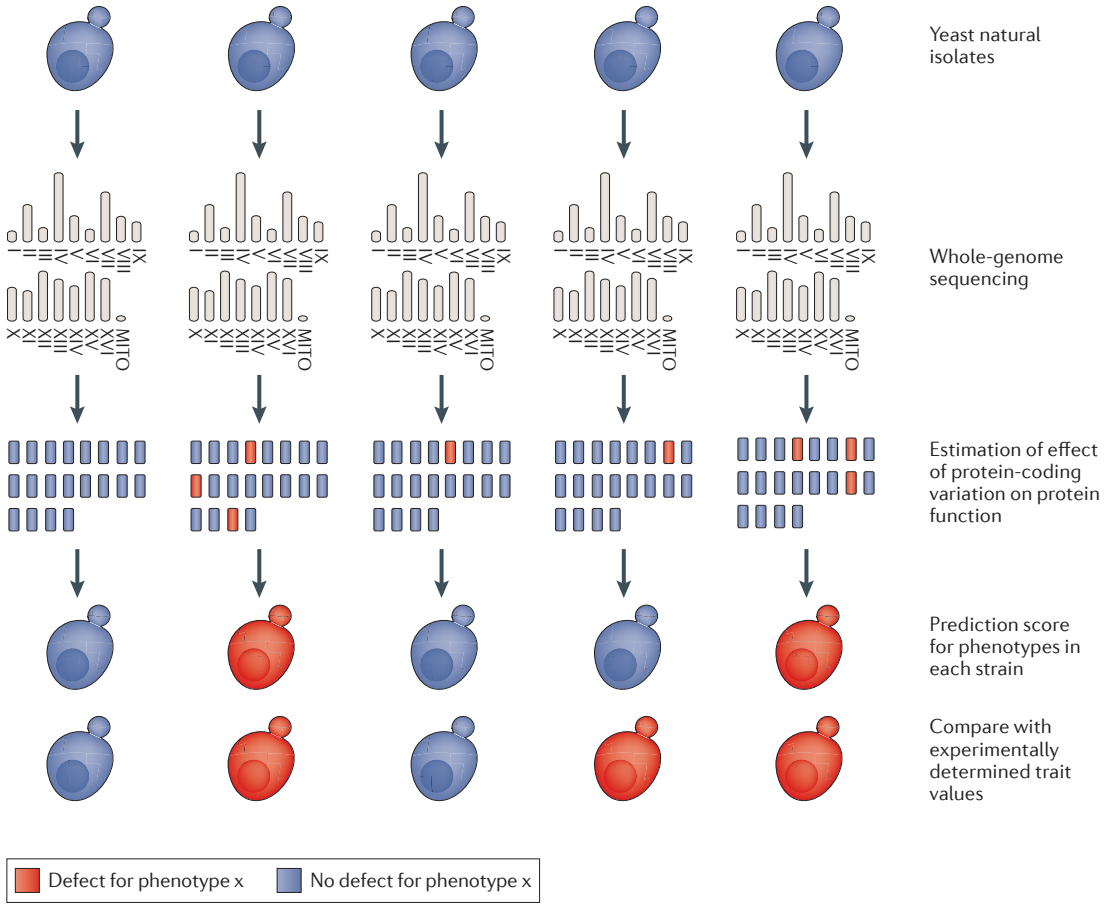

Figure 4 | **Whole-genome reverse genetics: making and evaluating phenotypic predictions from the genome sequences of individuals.** Model organisms can also be used to test methods for predicting phenotypic variation from whole-genome sequences. For example, individual yeast strains from around the world could be subjected to whole-genome sequencing and then diverse methods used to predict the genetic perturbations to different genes, pathways and processes believed to contribute to different phenotypic traits. In the example, the effects of mutations on individuals proteins in a set previously reported to influence a trait are estimated, and proteins predicted to have altered functions are indicated in red. The total perturbation in the gene set is then used to predict whether each individual will be affected for that trait (red) or not (blue) relative to a reference laboratory strain. These computational predictions can then be compared to the actual phenotypic variation quantified in laboratory experiments. Future studies could also assess the influence of non-protein-coding variation and changes in gene expression or copy number and could consider more complex models, epistatic interactions and non-homozygous genomes.

evaluated by their lack of clustering in an integrated network[86]. This study is, however, only a first step, and the approach that it proposes could be more widely used to evaluate alternative methods for predicting phenotypic variation from whole-genome sequences. For example, information on regulatory regions, gene expression measurements and epistatic interactions could all be incorporated into more sophisticated models. Indeed, it could be envisaged that multiple different groups could make phenotypic predictions from a common set of individual genome sequences, and the performance of these predictions could then be determined by independent experimental evaluation. Such cycles of prediction and independent evaluation in model organisms might be one way to improve methods for predicting phenotypic variation from whole-genome sequences.

Additional influences on trait variation

As noted above, many disease-associated mutations are incompletely penetrant (that is, not all individuals carrying a mutation develop a disease) or have variable expressivity (that is, individuals differ in the severity of disease). These phenomena are often assumed to be caused by either additive or epistatic interactions with other genetic variants in a genome or by interactions with environmental risk factors. However, incomplete penetrance and variable expressivity are common even in identical twins[5] and in inbred model organisms, such as mice and *C. elegans*, that can be raised in highly controlled environments[3]. For example, inbred rodent strains still show substantial variation in body weight even when the environment is tightly controlled[89]. What are the causes of this variation in 'genome outcome', even when the environment is controlled? The contributions

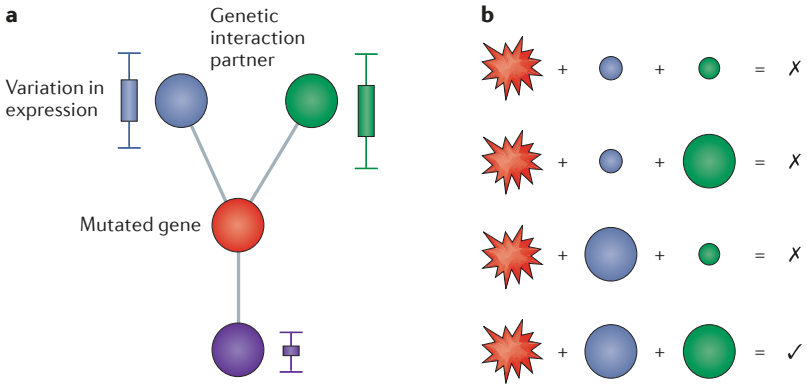

Figure 5 | **Epistasis with only a single mutation: inter-individual variation in the expression of genetic interaction partners contributes to incomplete penetrance.** **a** | If the effects of a mutation in one gene (red) can be enhanced by mutations in three other genes (blue, green and purple), then variation in the expression levels of these three genes may also enhance this mutation. In the example, two of these genes (blue and green) show substantial expression variation during early development, as indicated by the box plots next to each gene. **b** | In one possible scenario, the effects of the mutation in the red gene may be detrimental if the expression of either of the genetic interaction partners (blue and green) is low during development. By contrast, the mutation may have no effect when the expression of both of these genes is high. Such 'epigenetic' inter-individual variation in the expression levels of genetic interaction partners may be a common cause of incomplete penetrance and of the variable expressivity of mutations. This variation in gene expression can be stochastic, environmentally induced or perhaps subject to parental control.

to this are only just starting to be investigated and so are introduced in the final section of this Review.

Epigenetic epistasis. Recent studies have provided insights into this question, linking phenotypic variation to inter-individual differences in gene expression[3]. For example, in the bacterium *Bacillus subtilis*, variation in the outcome of a mutation in a gene affecting sporulation was partially accounted for by variation in the expression level of that gene[90]. A second study in *C. elegans* examined how genes vary in expression downstream of an incompletely penetrant mutation, showing that in the presence of an upstream mutation several downstream genes are not expressed, and levels of the remaining (but highly variable) active downstream gene are sometimes insufficient to activate the final gene in the regulatory cascade[91].

A third study proposed a more general model for incomplete penetrance, suggesting that it is variation in genetic interaction partners that underlies variation in the outcome of a mutation[92]. The logic of this model is the following: if the effects of a mutation in a gene are known to be influenced by mutations in a second gene, then non-genetic variation in the activity of this second gene might also influence the outcome of the mutation (FIG. 5). Thus, knowing the genetic interaction partners of a gene, it is possible to predict which genes, if they have sufficient inter-individual expression variation, might underlie incomplete penetrance or expressivity. These interactions between mutations and expression variants can be referred to as 'epigenetic' interactions or epigenetic epistasis.

Isogenic
Lacking genetic variation. Some laboratory animals, such as *Caenorhabditis elegans* and mice, are inbred and so siblings have identical genome sequences except for *de novo* mutations arising in each generation.

Epigenetic epistatic interactions can occur both because of variation in the expression of specific genetic interaction partners and because of variation in more promiscuous genetic interaction hubs. For example, during the early embryonic development of *C. elegans*, substantial inter-individual variation in the expression of the heat-shock protein 90 (HSP90) chaperone DAF-21 is observed, and this variation partially predicted variation in the outcome of a chaperone-dependent mutation[92]. As noted above, systematic screens in model organisms have shown that most genes have many potential genetic interaction partners, meaning that variation in the expression or activity of multiple genes could have an impact on the outcome of a particular mutation. By simultaneously quantifying the expression of two genetic interaction partners — a partially redundant paralogue of the mutated gene and the promiscuous genetic hub HSP90 — during early embryonic development, it was possible to predict more accurately the outcome of an inherited mutation in *C. elegans*[92].

Variation in how an organism responds to an environmental challenge can also underlie variation in the outcome of mutations. For example, following a mild heat stress, not all *C. elegans* larvae respond similarly, and longer signal duration in some individuals is linked to stronger induction of target genes, such as chaperones[85]. The ability to buffer the effects of inherited mutations was also stronger in individuals inducing a more substantial stress response[85]. Thus, gene–environment interactions can also vary among isogenic individuals, and inter-individual differences in gene expression lead to stronger or weaker genotype–environment interactions.

To date, inter-individual variation in mutation outcome has primarily been studied at the level of variation in gene expression. However, it is likely that variation at other scales could also have an impact on phenotypic variation, such as variation in cell contacts or mechanical stresses during development or variation in protein aggregation later in life. If such variation is temporally stable — for example, in the form of epigenetic inheritance through mitotic divisions that is transmitted by changes in chromatin or gene circuits — then early stochastic events may also influence how an organism later responds to a perturbation such as infection or diet. Thus, adult traits such as diabetes may also partially trace back to embryonic events[93,94]. Epigenetically stable stochastic variation could also be important in the initiation of tumour clones: inter-individual cellular variation that provides a growth advantage will be selected and, if it is semi-stably mitotically inherited, may lead to cancer.

Finally, although appreciated as an important influence in cancer, somatic mutations acquired during an individual's development may also contribute more generally to trait variation.

Parental influences on phenotypic variation. Two additional influences on phenotypic variation that are often overlooked and that are still quite poorly understood at the molecular level are non-inherited genetic variation (that is, the genotypes of parents) and transgenerational

influences of the environment (for example, alterations in maternal provisioning of embryos).

Examples of non-inherited genetic variation that influence phenotypic variation are quite common in model organisms; examples include maternal-effect mutations that affect *Drosophila melanogaster* development and a natural polymorphism with a paternal effect in *C. elegans* that causes lethality in particular zygotic genotypes[95]. Another example is the modification of tumour growth in *D. melanogaster*: individual flies that inherit a hyperactive JAK kinase develop tumours if their parents carry mutations in several different regulators[96], and the authors propose that the JAK kinase antagonizes the erasure of parentally derived epigenetic markings induced by these mutations. These examples highlight the potential for the consequences of inherited mutations to be influenced by genetic variation that was present in parents but not inherited by their progeny.

In vertebrates and invertebrates, there are also diverse examples of how changes in the parental environment can influence phenotypic variation in the next generation[3]. In some cases, these changes are proximal: for example, in *C. elegans*, osmotic stress triggers increased deposition of glycerol into oocytes with the result that these oocytes are better protected from osmotic stress but are more susceptible to hypoxia[97]. In other cases, the effects of parental environment are longer lasting: for example, certain strains of male rat that are fed a chronic high-fat diet are more likely to have female offspring with pancreatic β-cell dysfunction[98], and male mice that are fed a low-protein diet have offspring with altered metabolic gene expression in their livers[99]. Such parental influences may, in some cases, also be transmitted for multiple generations[100]. Indeed, in humans there is epidemiological evidence that the environment experienced by one generation might influence the phenotypes of subsequent generations[101].

In most cases, how the maternal or paternal environment influences phenotypic variation in offspring is not understood at the molecular level, and dissecting these molecular mechanisms is a key challenge for the field. However, the phenomenon of imprinting in mammals — whereby either the maternal or paternal copy of a gene is silenced — clearly shows that there is the potential for the propagation of specific epigenetic information from the germ line to a zygote[102]. Moreover, the establishment of transgenerationally inherited gene silencing by small RNA pathways in *C. elegans* also highlights the potential for specific epigenetic information to be transferred across generations[103–105], as does the transgenerational propagation of the effects of genetic perturbations that affect lifespan[106].

Future challenges

In this Review, I have highlighted some recent work in model organisms that is relevant to the problem of making accurate phenotypic predictions in individual humans. Of course, there remain many important challenges, and I discuss a few of these here.

First, with respect to the problem of linking genes to traits, one major goal should be the creation of gene–phenotype maps for vertebrates similar to those that have been produced for model organisms. These model-organism maps — produced using both RNAi and genetics and generated using both cells and whole animals — would serve as a framework for human genetic studies. One technical challenge is the 'sign problem' of positive and negative regulators: to make accurate genetic predictions, all of the genes that influence a human disease need to be distinguished, plus their direction of action and any interactions. Similarly, the ability to distinguish loss-of-function from gain-of-function or change-of-function mutations has received little attention but is important for making integrated predictions across genomes. Haploinsufficiency[107] and dominance[20] have also received little attention in the context of genome-scale predictions, and the mechanisms that cause small increases in the dosage of some genes to have phenotypic consequences are also not clear and warrant further investigation. In addition, predicting the effects of variants in non-coding regions is still an open challenge, and although improved genome annotation and cross-species analysis should facilitate this, extensive sets of phenotypically relevant mutations are also likely to be required.

Second, with respect to epistasis, a major challenge is to understand the molecular mechanisms that underlie most epistatic interactions[50]. Although epistatic interactions often fall 'between pathways' (or between modules)[70–72], it is not clear why different modules interact or how these interactions can be predicted *de novo*. Other issues that remain to be systematically investigated are the importance of epistasis involving weak alleles, rather than null alleles or those that strongly reduce function, and the extent of epistasis between heterozygous mutations. Moreover, epistasis screens with gain-of-function mutations have been limited to a small number of overexpression screens[108], and this is another area that deserves more attention. However, the development of computational methods to predict epistatic interactions in human disease genetics is perhaps the most pressing challenge.

Third, with respect to variation in non-protein-coding regions of the genome, substantial efforts are required to identify systematically the regulatory regions of each gene and to build computational models of when polymorphisms in these regions affect expression and phenotypic traits.

Fourth, with respect to gene–environment interactions and whole-genome reverse genetics, a key challenge will be to make predictions in more complex scenarios, such as from heterozygous genomes, for higher model organisms and for traits for which there is less complete knowledge about the relevant genes. An additional challenge will be to incorporate other kinds of genetic variation, such as copy number changes and variants in non-coding regions, as well as gain-of-function mutations. Of course, an important question is also whether considering epistatic interactions between variants will be required to make more accurate predictions and how best to predict these. Moreover, will complex and dynamic models, such as those involving metabolic networks or regulatory interactions, or those focused on a particular pathway or process be required for more accurate predictions?

Haploinsufficiency
A gene is haploinsufficient if removal of one of the two copies in a diploid organism has a detectable effect on fitness or a phenotype.

Dominance
The extent to which one allele of a gene exerts its effects irrespective of a second allele in diploid organisms. Complete dominance implies that the heterozygote has a phenotype that is indistinguishable from that of the dominant homozygote. Overdominance implies that the phenotype of the heterozygote lies outside the range of both homozygote parents.

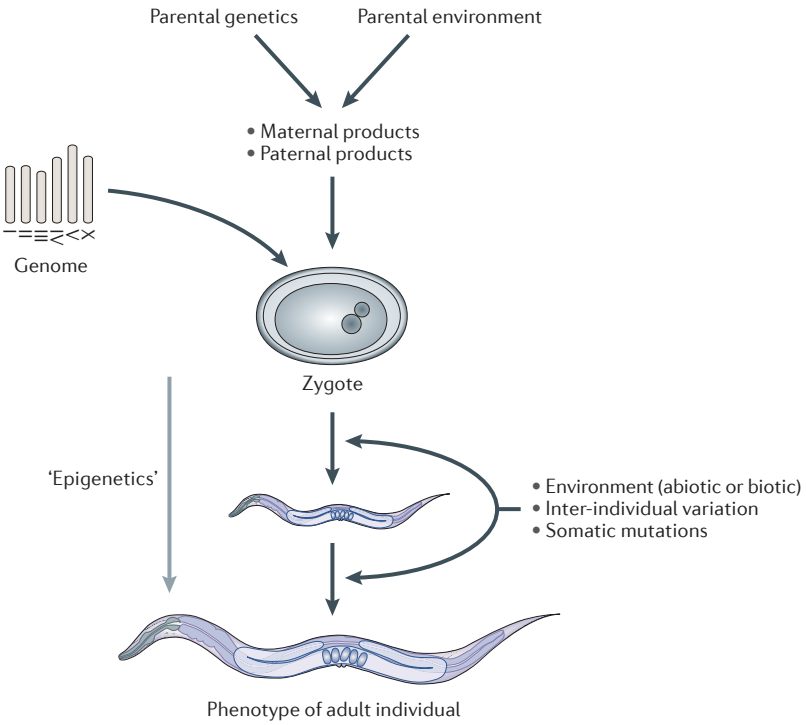

Figure 6 | **Sources of phenotypic variance in individuals.** A complete understanding of phenotypic variation in individuals will require an understanding of the contributions of multiple sources of variance and how they interact. For example, as illustrated here for *Caenorhabditis elegans*, phenotypic traits may be influenced by variation in an individual's genome, by maternal or paternal products contributed to the zygote that may be influenced by parental genotype or parental environment, by somatic mutations, by inter-individual stochastic variation and by both biotic (for example, pathogens, commensal microbiota or parental behaviour) and abiotic (for example, diet or temperature) environmental factors experienced at different life stages. The extent to which early variation influences later phenotypic variation and how early variation is propagated ('epigenetics') are also important open questions.

Fifth, with respect to variation in the outcome of mutations among isogenic individuals, it will be necessary to test whether the models developed in invertebrates also apply to vertebrate systems. Does variation during embryonic development have an impact on the developmental outcome of mutations and on adult phenotypes in other species? In addition, the extent to which environmentally triggered responses and parental genetics can alter phenotypic variation in subsequent generations warrants much more investigation, and the molecular mechanisms that underlie these transgenerational effects need to be elucidated. Ultimately, it will be necessary to understand how all of the various influences on phenotypic traits throughout an individual's life interact to determine their final characteristics (FIG. 6).

Concluding remarks

I have attempted to highlight here how model organisms are being used to develop and to evaluate methods to link genetic variation to phenotypic variation more comprehensively and also to understand why accurate phenotypic predictions may, for many traits and diseases, never be possible from genome sequencing alone. Rather, the work from model organisms reminds us that to make accurate predictions at the level of individuals, it will be necessary to combine genetic information with appropriate *in vivo* measurements of physiological states and other 'intermediate phenotypes', such as gene expression, protein and metabolite levels or other functional assays that capture additional influences on trait variation[109]. As highlighted above, there still remain many open questions, and model organisms will continue to provide an intellectual framework, directly transferable biological knowledge and practical computational methods that can be applied to human genetics.

1. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
2. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
3. Burga, A. & Lehner, B. Beyond genotype to phenotype: why the phenotype of an individual cannot always be predicted from their genome sequence and the environment that they experience. *FEBS J.* **279**, 3765–3775 (2012).
4. Clayton, D. G. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLoS Genet.* **5**, e1000540 (2009).
5. Roberts, N. J. *et al.* The predictive capacity of personal genome sequencing. *Sci. Transl. Med.* **4**, 133ra58 (2012).
This study provides estimates of the maximum ability of whole-genome sequencing to predict clinically useful risk information for 24 diseases on the basis of analyses of monzygotic twin pairs.
6. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
7. Kim, D. U. *et al.* Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nature Biotech.* **28**, 617–623 (2010).
8. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
9. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
This is the first genome-wide analysis of the effects of gene function inhibition in an animal.
10. Dietzl, G. *et al.* A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
11. Hobert, O. The impact of whole genome sequencing on model system genetics: get ready for the ride. *Genetics* **184**, 317–319 (2010).
12. Ehrenreich, I. M. *et al.* Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**, 1039–1042 (2010).
13. Ehrenreich, I. M. *et al.* Genetic architecture of highly complex chemical resistance traits across four yeast strains. *PLoS Genet.* **8**, e1002570 (2012).
This paper describes the detection of more than 800 loci that influence resistance to 13 chemicals in all 6 pairwise crosses of four yeast strains, using extremely large pools of segregants.
14. Liti, G. & Louis, E. J. Advances in quantitative trait analysis in yeast. *PLoS Genet.* **8**, e1002912 (2012).
15. Parts, L. *et al.* Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* **21**, 1131–1138 (2011).
16. Swinnen, S. *et al.* Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence analysis. *Genome Res.* **22**, 975–984 (2012).
17. Dudley, A. M., Janse, D. M., Tanay, A., Shamir, R. & Church, G. M. A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol. Syst. Biol.* **1**, 2005.0001 (2005).
18. Brown, J. A. *et al.* Global analysis of gene function in yeast by quantitative phenotypic profiling. *Mol. Syst. Biol.* **2**, 2006.0001 (2006).
19. Hillenmeyer, M. E. *et al.* The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–365 (2008).
20. Wright, S. Physiological and evolutionary theories of dominance. *Am. Nat.* **68**, 24–53 (1934).
21. McGary, K. L. *et al.* Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc. Natl Acad. Sci. USA* **107**, 6544–6549 (2010).
This paper describes the systematic identification of 'phenologues', which are phenotypes in different species that are linked because they are affected by overlapping sets of genes.
22. Cha, H. J. *et al.* Evolutionarily repurposed networks reveal the well-known antifungal drug thiabendazole to be a novel vascular disrupting agent. *PLoS Biol.* **10**, e1001379 (2012).
23. Fraser, A. G. & Marcotte, E. M. A probabilistic view of gene function. *Nature Genet.* **36**, 559–564 (2004).
24. Lehner, B. & Lee, I. Network-guided genetic screening: building, testing and using gene networks to predict gene function. *Brief. Funct. Genom. Proteom.* **7**, 217–227 (2008).
25. Lee, I., Date, S. V., Adai, A. T. & Marcotte, E. M. A probabilistic functional network of yeast genes. *Science* **306**, 1555–1558 (2004).

26. Troyanskaya, O. G., Dolinski, K., Owen, A. B., Altman, R. B. & Botstein, D. A. Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*). *Proc. Natl Acad. Sci. USA* **100**, 8348–8353 (2003).
27. Rhodes, D. R. *et al.* Probabilistic model of the human protein–protein interaction network. *Nature Biotech.* **23**, 951–959 (2005).
28. McGary, K. L., Lee, I. & Marcotte, E. M. Broad network-based predictability of *Saccharomyces cerevisiae* gene loss-of-function phenotypes. *Genome Biol.* **8**, R258 (2007).
29. Lee, I. *et al.* A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*. *Nature Genet.* **40**, 181–188 (2008).
30. Lee, I., Li, Z. & Marcotte, E. M. An improved, bias-reduced probabilistic functional gene network of baker's yeast, *Saccharomyces cerevisiae*. *PLoS ONE* **2**, e988 (2007).
31. Li, Z. *et al.* Rational extension of the ribosome biogenesis pathway using network-guided genetics. *PLoS Biol.* **7**, e1000213 (2009).
32. Myers, C. L. *et al.* Discovery of biological networks from diverse functional genomic data. *Genome Biol.* **6**, R114 (2005).
33. Chikina, M. D., Huttenhower, C., Murphy, C. T. & Troyanskaya, O. G. Global prediction of tissue-specific gene expression and context-dependent gene networks in *Caenorhabditis elegans*. *PLoS Comput. Biol.* **5**, e1000417 (2009).
34. Pena-Castillo, L. *et al.* A critical assessment of *Mus musculus* gene function prediction using integrated genomic evidence. *Genome Biol.* **9** (Suppl. 1), S2 (2008).
35. Lee, I., Ambaru, B., Thakkar, P., Marcotte, E. M. & Rhee, S. Y. Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. *Nature Biotech.* **28**, 149–156 (2010).
36. Lee, I. *et al.* Genetic dissection of the biotic stress response using a genome-scale gene network for rice. *Proc. Natl Acad. Sci. USA* **108**, 18548–18553 (2011).
37. Lage, K. *et al.* A human phenome–interactome network of protein complexes implicated in genetic disorders. *Nature Biotech.* **25**, 309–316 (2007).
38. Lee, I., Blom, U. M., Wang, P. I., Shim, J. E. & Marcotte, E. M. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* **21**, 1109–1121 (2011).
39. Gillis, J. & Pavlidis, P. The impact of multifunctional genes on "guilt by association" analysis. *PLoS ONE* **6**, e17258 (2011).
40. Sopko, R. *et al.* Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319–330 (2006).
41. Vavouri, T., Semple, J. I., Garcia-Verdugo, R. & Lehner, B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* **138**, 198–208 (2009).
42. Birchler, J. A. & Veitia, R. A. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl Acad. Sci. USA* **109**, 14746–14753 (2012).
43. Moriya, H., Shimizu-Yoshida, Y. & Kitano, H. *In vivo* robustness analysis of cell division cycle genes in *Saccharomyces cerevisiae*. *PLoS Genet.* **2**, e111 (2006).
44. Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
45. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nature Rev. Genet.* **10**, 184–194 (2009).
46. Francesconi, M., Jelier, R. & Lehner, B. Integrated genome-scale prediction of detrimental mutations in transcription networks. *PLoS Genet.* **7**, e1002077 (2011).
47. Gertz, J., Siggia, E. D. & Cohen, B. A. Analysis of combinatorial *cis*-regulation in synthetic and genomic promoters. *Nature* **457**, 215–218 (2009).
48. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature Biotech.* **30**, 521–530 (2012).
49. Phillips, P. C. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev. Genet.* **9**, 855–867 (2008).
50. Lehner, B. Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27**, 323–331 (2011).
51. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* **109**, 1193–1198 (2012).
52. Lehner, B. Modelling genotype–phenotype relationships and human disease with genetic interaction networks. *J. Exp. Biol.* **210**, 1559–1566 (2007).
53. Drees, B. L. *et al.* Derivation of genetic interaction networks from quantitative phenotype data. *Genome Biol.* **6**, R38 (2005).
54. Phillips, P. C. The language of gene interaction. *Genetics* **149**, 1167–1171 (1998).
55. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
The most comprehensive analysis of epistatic interactions in any organism; the effects on growth are quantified for more than 5 million pairs of mutations in yeast.
56. Frost, A. *et al.* Functional repurposing revealed by comparing *S. pombe* and *S. cerevisiae* genetic interactions. *Cell* **149**, 1339–1352 (2012).
57. Ryan, C. J. *et al.* Hierarchical modularity and the evolution of genetic interactomes across species. *Mol. Cell* **46**, 691–704 (2012).
58. Lehner, B., Crombie, C., Tischler, J., Fortunato, A. & Fraser, A. G. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nature Genet.* **38**, 896–903 (2006).
59. Byrne, A. B. *et al.* A global analysis of genetic interactions in *Caenorhabditis elegans*. *J. Biol.* **6**, 8 (2007).
60. Horn, T. *et al.* Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nature Methods* **8**, 341–346 (2011).
61. Tong, A. H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
62. Gerke, J., Lorenz, K. & Cohen, B. Genetic interactions between transcription factors cause natural variation in yeast. *Science* **323**, 498–501 (2009).
63. Lorenz, K. & Cohen, B. A. Small- and large-effect quantitative trait locus interactions underlie variation in yeast sporulation efficiency. *Genetics* **192**, 1123–1132 (2012).
64. Brem, R. B. & Kruglyak, L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl Acad. Sci. USA* **102**, 1572–1577 (2005).
65. Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436**, 701–703 (2005).
This paper highlights the importance of epistatic interactions between natural variants that influence gene expression.
66. Dowell, R. D. *et al.* Genotype to phenotype: a complex problem. *Science* **328**, 469 (2010).
By constructing a gene deletion collection for a second laboratory strain of yeast, the authors identify >40 genes that are essential in this strain but not in a previously analysed strain. In most cases, this 'conditional' essentiality in one strain is due to variation in four or more different modifier loci.
67. Koch, E. N. *et al.* Conserved rules govern genetic interaction degree across species. *Genome Biol.* **13**, R57 (2012).
68. Wong, S. L. *et al.* Combining biological networks to predict genetic interactions. *Proc. Natl Acad. Sci. USA* **101**, 15682–15687 (2004).
69. Lee, I. *et al.* Predicting genetic modifier loci using functional gene networks. *Genome Res.* **20**, 1143–1153 (2010).
70. Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein networks. *Nature Biotech.* **23**, 561–566 (2005).
71. Ulitsky, I. & Shamir, R. Pathway redundancy and protein essentiality revealed in the *Saccharomyces cerevisiae* interaction networks. *Mol. Syst. Biol.* **3**, 104 (2007).
72. Bellay, J. *et al.* Putting genetic interactions in context through a global modular decomposition. *Genome Res.* **21**, 1375–1387 (2011).
73. Hess, D. C. *et al.* Computationally driven, quantitative experiments discover genes required for mitochondrial biogenesis. *PLoS Genet.* **5**, e1000407 (2009).
74. Gerke, J., Lorenz, K., Ramnarine, S. & Cohen, B. Gene–environment interactions at nucleotide resolution. *PLoS Genet.* **6**, e1001144 (2010).
75. St Onge, R. P. *et al.* Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nature Genet.* **39**, 199–206 (2007).
76. Bandyopadhyay, S., Kelley, R., Krogan, N. J. & Ideker, T. Functional maps of protein complexes from quantitative genetic interaction data. *PLoS Comput. Biol.* **4**, e1000065 (2008).
77. Harrison, R., Papp, B., Pal, C., Oliver, S. G. & Delneri, D. Plasticity of genetic interactions in metabolic networks of yeast. *Proc. Natl Acad. Sci. USA* **104**, 2307–2312 (2007).
78. Dixon, S. J. *et al.* Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proc. Natl Acad. Sci. USA* **105**, 16653–16658 (2008).
79. Tischler, J., Lehner, B. & Fraser, A. G. Evolutionary plasticity of genetic interaction networks. *Nature Genet.* **40**, 390–391 (2008).
80. Roguev, A. *et al.* Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* **322**, 405–410 (2008).
81. Lindquist, S. Protein folding sculpting evolutionary change. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 103–108 (2009).
82. Zhao, R. *et al.* Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* **120**, 715–727 (2005).
83. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
84. Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
85. Casanueva, M. O., Burga, A. & Lehner, B. Fitness trade-offs and environmentally induced mutation buffering in isogenic *C. elegans*. *Science* **335**, 82–85 (2012).
86. Jelier, R., Semple, J. I., Garcia-Verdugo, R. & Lehner, B. Predicting phenotypic variation in yeast from individual genome sequences. *Nature Genet.* **43**, 1270–1274 (2011).
This paper reports the application of whole-genome reverse genetics: phenotypic predictions are made from the complete genome sequences of yeast strains and the accuracy of these predictions are evaluated by experimentation. Predictions are generally good, even for genetically complex traits, when the genes linked to the trait are evaluated as reliable using a gene network.
87. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).
88. Baker, M. Functional genomics: the changes that count. *Nature* **482**, 257–262 (2012).
89. Gartner, K. A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? *Lab. Anim.* **24**, 71–77 (1990).
90. Eldar, A. *et al.* Partial penetrance facilitates developmental evolution in bacteria. *Nature* **460**, 510–514 (2009).
91. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918 (2010).
92. Burga, A., Casanueva, M. O. & Lehner, B. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature* **480**, 250–253 (2011).
Whether an inherited mutation affects genetically identical individuals or not is predicted by inter-individual variation in the expression of a specific and a promiscuous genetic interaction partner during early embryonic development.
93. Hales, C. N. & Barker, D. J. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia* **35**, 595–601 (1992).
94. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nature Med.* **17**, 448–453 (2011).

95. Seidel, H. S., Rockman, M. V. & Kruglyak, L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* **319**, 589–594 (2008).
96. Xing, Y. *et al.* Evidence for transgenerational transmission of epigenetic tumor susceptibility in *Drosophila*. *PLoS Genet.* **3**, 1598–1606 (2007).
97. Frazier, H. N. & Roth, M. B. Adaptive sugar provisioning controls survival of *C. elegans* embryos in adverse environments. *Curr. Biol.* **19**, 859–863 (2009).
98. Ng, S. F. *et al.* Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring. *Nature* **467**, 963–966 (2010).
99. Carone, B. R. *et al.* Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084–1096 (2010).
100. Jablonka, E. & Raz, G. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q. Rev. Biol.* **84**, 131–176 (2009).
101. Painter, R. C. *et al.* Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life. *BJOG* **115**, 1243–1249 (2008).
102. Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. *Nature Rev. Genet.* **12**, 565–575 (2011).
103. Ashe, A. *et al.* piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell* **150**, 88–99 (2012).
104. Shirayama, M. *et al.* piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell* **150**, 65–77 (2012).
105. Buckley, B. A. *et al.* A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature* **489**, 447–451 (2012).
References 103–105 establish that piRNA-triggered gene silencing is stably transmitted across many generations in *C. elegans*.
106. Greer, E. L. *et al.* Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nature* **479**, 365–371 (2011).
107. Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* **6**, e1001154 (2010).
108. Liu, C., van Dyk, D., Li, Y., Andrews, B. & Rao, H. A genome-wide synthetic dosage lethality screen reveals multiple pathways that require the functioning of ubiquitin-binding proteins Rad23 and Dsk2. *BMC Biol.* **7**, 75 (2009).
109. Burga, A. & Lehner, B. Predicting phenotypes from genotypes, phenotypes and a combination of the two. *Curr. Opin. Biotech.* (in the press).

Acknowledgements
Our research is funded by the European Research Council (ERC), MINECO Plan Nacional grants BFU2008-00365 and BFU2011-26206, ERASysBio+ ERANET project EUI2009-04059 GRAPPLE, the European Molecular Biology Organization (EMBO) Young Investigator Program, EU Framework 7 project 277899 4DCellFate and the EMBL/CRG Systems Biology Program.

Competing interests statement
The author declares no competing financial interests.

FURTHER INFORMATION
Author's homepage: http://www.crg.eu/ben_lehner

ALL LINKS ARE ACTIVE IN THE ONLINE PDF