The Journal of Experimental Biology 210, 1559-1566
Published by The Company of Biologists 2007
doi:10.1242/jeb.002311

Modelling genotype–phenotype relationships and human disease with genetic interaction networks

Ben Lehner

EMBL/CRG Systems Biology Unit, Centre for Genomic Regulation (CRG), UPF, C/Dr Aiguader 88, Barcelona 08003, Spain

e-mail: ben.lehner@crg.es

Accepted 16 January 2007

Summary

Probably all heritable traits, including disease susceptibility, are affected by interactions between mutations in multiple genes. We understand little, however, about how genes interact to produce phenotypes, and there is little power to detect interactions between genes in human population studies. An alternative approach towards understanding how mutations combine to produce phenotypes is to construct systematic genetic interaction networks in model organisms. Here I describe the methods that are being used to map genetic interactions in yeast and *C. elegans*, and the insights that these networks provide for human disease. I also discuss the mechanistic interpretation of genetic interaction networks, how genetic interactions can be used to understand gene function, and methods that have been developed to predict genetic interactions on a genome-wide scale.

Glossary available online at
http://jeb.biologists.org/cgi/content/full/210/9/1559/DC1

Key words: genetic interactions, networks, systems biology.

Introduction

The relationship between the genotype of an organism and its phenotype is not a simple one-to-one mapping between genes and phenotypes. Rather phenotypes result from the interactions between the products of many different genes. The complexity of this relationship is well illustrated by the genetics of disease in humans: probably all hereditary diseases in humans are genetically complex, resulting not from mutations in a single gene, but from the combination of mutations in multiple different genes (Badano and Katsanis, 2002). For example even in the case of the simple ‘Mendelian’ disease, cystic fibrosis, it is not possible to predict the clinical phenotype of a patient based solely on knowledge of the exact mutation in the ‘Cystic Fibrosis Gene’, CFTR. In fact at least seven different modifier genes have been described that alter the clinical phenotype of this genetically simple disease (Badano and Katsanis, 2002).

Most hereditary diseases are genetically much more complex than cystic fibrosis; although an increasing number of genes have been identified as mutated in common pathologies such as cardiovascular disease, cancer, diabetes and neurodegenerative diseases, these mutations only account for a small proportion of the total genetic predisposition to these conditions (Badano and Katsanis, 2002). One reason why causal mutations have proved so difficult to identify may be the problem of synthetic interactions between genes: mutations that have little effect on disease phenotypes alone can have strong synthetic effects when combined (Hartman et al., 2001). Indeed in most linkage or association studies there is insufficient statistical power to identify these interactions between genes (Badano and Katsanis, 2002). Therefore the extent and importance of genetic interactions in human disease remains largely unknown.

An alternative approach to understand how genes interact to produce phenotypes is to identify genetic interactions between mutations in model organisms (Hartman et al., 2001). The idea of this approach is to take a simple phenotype (typically the simplest, viability) and to identify comprehensively how combinations of mutations in genes can affect this phenotype. Although there are many important types of aggravating and alleviating genetic interactions that can occur between genes (Drees et al., 2005), to date most work has concentrated on synthetic lethal interactions. A synthetic (or synergistic) lethal interaction is formally defined when the survival resulting from combining mutations in two genes is less than the product of the survival resulting from each mutation individually (Drees et al., 2005). Most commonly synthetic lethal interactions are identified experimentally when the

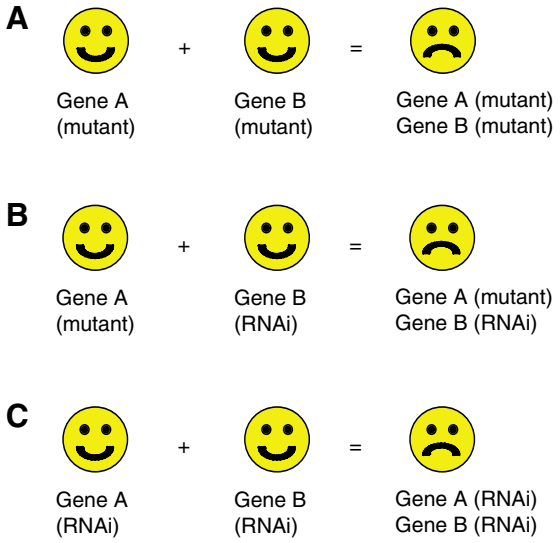

Fig. 1. Synthetic genetic interactions. A synthetic lethal interaction between two genes is defined when the survival of the combined mutation is less than the product of the survival of the two single mutations. In yeast, genetic interactions are defined by combining mutant strains using systematic mating protocols (A), and synthetic lethal or sick phenotypes are defined where a double mutant strain displays a phenotype that is not seen with either single mutant strain. In *C. elegans*, genetic interations are defined by combining genetic mutations with RNAi to target a second gene (B), or by using combinatorial RNAi to target two genes simultaneously (C) (Tischler et al., 2006). Synthetic aggravating phenotypes can be similarly defined for other phenotypes such as sterility or growth (Lehner et al., 2006b), and many more possible combinations of aggravating or alleviating interactions are also possible (Drees et al., 2005).

combination of mutations in two non-essential genes produces a lethal phenotype (Fig. 1).

Mapping genetic interaction networks in yeast

Currently the most extensively mapped genetic interaction network exists for the budding yeast *Saccharomyces cerevisiae* (Ooi et al., 2006; Pan et al., 2006; Schuldiner et al., 2005; Tong et al., 2001; Tong et al., 2004). Two general strategies have been devised that make use of the library of yeast gene deletion strains (Giaever et al., 2002) to identify genetic interactions systematically in this haploid organism. In the first approach, double mutants are constructed and assayed for viability in parallel by mating a yeast strain carrying a query mutation to the complete library of viable deletion strains in an arrayed format [‘synthetic genetic array’ analysis, SGA (Tong et al., 2001)]. In contrast, in the second approach double mutants are constructed and assayed as a single pool, and the relative growth rates of each strain are determined using DNA microarrays that can detect the presence of each strain in the pool [‘synthetic lethal analysis by microarray’, SLAM (Ooi et al., 2003)]. I will first outline these two methods, and then will discuss the genetic interaction networks that have been produced using them.

SGA

In the SGA approach (Tong et al., 2001), a haploid yeast strain carrying a ‘query’ mutation in a gene of interest is mated to a library of yeast deletion strains in an arrayed format using replicating tools and robotics. The diploid yeast are then sporulated, and double mutant haploid progeny are selected using a cleverly designed reporter construct (the ‘SGA reporter’, *can1Δ::MFA1pr-HIS3*, that is present in the *MATalpha* query strain, but in the absence of histidine only allows growth of haploids of mating type *MATa*, i.e. only the double mutant progeny). Synthetic sick or nonviable double mutants are identified by weakly growing or absent double mutant colonies, and their identity is determined by their position on the array. Potential synthetic lethal or sick interactions are then individually confirmed by tetrad or random spore analysis (Tong et al., 2001).

SLAM

In the SLAM approach, a query mutation is introduced into the pool of haploid deletion strains by direct integrative transformation (Ooi et al., 2003). The double mutants are then grown competitively in a single vessel, and non-growing or slow growing double mutant strains are identified using microarrays. This approach is possible because of the two ‘barcode’ DNA sequences that uniquely identify each deletion strain. These barcodes allow the deletion strains that are present in a pool to be individually identified by hybridisation of genomic DNA to a microarray containing sequences complementary to each of the barcodes. In contrast to the qualitative SGA approach, in the SLAM procedure the definition of a synthetic interaction depends upon a quantitative cut-off in hybridisation intensity to identify slow growing or absent strains. A modification of the SLAM procedure uses heterozygous diploid deletion strains as a starting point [‘diploid-based SLAM’ or dSLAM (Pan et al., 2004)]. Maintaining the deletion strains as heterozygous diploids protects them from selection for compensatory or reversion mutations that overcome fitness defects, so reducing the false negative rate of the approach. As a result, the dSLAM approach probably has a lower false negative rate than SGA (Tong et al., 2004). One disadvantage of the dSLAM approach is that some barcode tags have low hybridisation-signal intensities (for example because of mutations in the tag sequences) with the result that there is little reliable information for some genes (Eason et al., 2004). However this limitation has been addressed by redesigning the microarrays used to detect the barcodes (Pierce et al., 2006; Yuan et al., 2005).

Both the SGA and dSLAM approaches have been used to construct extensive genetic interaction networks in yeast. Using the SGA approach, Tong et al. screened 132 query strains (carrying mutations in genes with diverse functions in cell polarity, cell wall biosynthesis, chromosome segregation and DNA synthesis and repair) against the complete library of ~4700 viable haploid deletion strains, and identified a total of 2012 synthetic lethal and 2113 synthetic sick interactions

involving ~1000 genes (Tong et al., 2004). Both deletions of non-essential genes and point mutations in essential genes were used as query genes, and synthetic lethal interactions were detected for 80% of query strains, with a mean of 34 interactions per query gene (and a range of 1–146 interactions per gene). Using the dSLAM approach Pan et al. screened 74 query strains known to function in DNA replication and repair against the same deletion library and identified a total of 4956 synthetic fitness or lethality defects involving 875 genes (Pan et al., 2006). Over 91% of these interactions were entirely novel (Pan et al., 2006).

Mapping genetic interactions for essential genes

Both the Tong et al. and Pan et al. studies screened query strains against the ~4700 viable yeast deletion strains, so interactions with the ~1000 essential genes in the yeast genome could not be detected. Two approaches have been developed to identify genetic interactions with essential genes (Davierwala et al., 2005; Schuldiner et al., 2005). In the first approach, Davierwala et al. constructed a library of yeast strains that carry promoter-replacement alleles. These alleles allow the expression of each gene to be switched off by the addition of the small molecule doxycycline to the media (the 'tet-off' system). Addition of intermediate levels of doxycycline can therefore be used to reduce the expression of each essential gene, so producing hypomorphic (reduction in function) alleles of each gene. The authors created a library consisting of promoter replacement alleles for 575 essential genes (representing about half of the total number of essential genes) and screened it against 30 query strains that were either conditional alleles of essential genes or deletions of non-essential genes, identifying a total of 567 interactions. Interestingly the mean number of interactions detected for each essential gene was about sixfold more than for non-essential genes (Davierwala et al., 2005).

The second strategy that has been used to identify interactions for essential genes is to generate hypomorphic alleles by replacing the 3′UTR of each gene with an antibiotic resistance cassette ['decreased abundance by mRNA perturbation', DAmP (Schuldiner et al., 2005)]. In this approach, the antibiotic resistance cassette serves to destabilise the expression of an mRNA, so reducing the expression of each essential gene. The DAmP approach was used to identify genetic interactions between genes that function in the early secretory pathway (Schuldiner et al., 2005) (see below).

In theory, all of these methods for identifying interactions between pairs of genes could also be used to identify higher order interactions between more than two genes. For example, Tong et al. also used SGA to screen two different double mutant strains for interactions with a third gene to identify trigenic interactions (Tong et al., 2004). The authors identified a total of 171 and 156 interactions in these screens, although only 4 and 29 of the interactions were attributable to a triple mutant effect (the rest were also seen in one of the three possible double mutant combinations alone). However, because there are ~2000-fold more possible gene triplets than gene pairs in the *S. cerevisiae* genome, the total number of trigenic synthetic lethal interactions may still be greater than the number of digenic interactions.

Quantitative genetic interaction screens

The SGA and dSLAM approaches have been used to identify synthetic lethal and sick phenotypes on a genomic-scale. However there are many other classes of interactions that can occur between genes, and approaches have also been devised to begin to identify these interactions systematically (Collins et al., 2006; Drees et al., 2005; Hartman and Tippery, 2004; Schuldiner et al., 2005). For example, Schuldiner et al. used digital imaging to quantify the growth of yeast colonies such that they could measure both aggravating and alleviating interactions between 424 yeast genes that function in the early secretory pathway (Schuldiner et al., 2005). Rather than categorizing the observed interactions into different types of genetic interaction, they used a continuous score to describe the strength of an interaction and to cluster genes according to their interaction profiles. The authors demonstrated that using quantitative measurements of interaction strength helped to identify modules of genes that share precise molecular functions (Schuldiner et al., 2005).

Mapping genetic interaction networks in *C. elegans*

These studies in *S. cerevisiae* have provided unprecedented insight into the extent and properties of genetic interaction networks. However, *S. cerevisiae* is a unicellular yeast that does not contain many of the genes and pathways that are present in multicellular organisms. In particular, many of the signalling pathways implicated in human diseases such as cancer are not encoded in the genome of yeast. Therefore to understand how genes will interact in humans, it is essential that we should also systematically identify genetic interactions in a multicellular animal.

Systematically identifying genetic interactions by crossing mutant strains is not logistically practical in multicellular organisms – comprehensive collections of deletion strains are not available, and the diploidy of these organisms requires cumbersome multigenerational mating and selection screens to be used. An alternative approach for identifying genetic interactions in metazoans is to use RNA interference (RNAi) to inhibit gene expression (Baugh et al., 2005; Holway et al., 2005; Lehner et al., 2006a; Lehner et al., 2006b; Lehner et al., 2006c; Suzuki and Han, 2006; Tischler et al., 2006; van Haaften et al., 2004). Here either a genetic mutant is combined with RNAi against a second gene (Fig. 1B), or RNAi can be used to inhibit the expression of two genes simultaneously (Fig. 1C) (Tischler et al., 2006). One advantage of using RNAi compared to deletion strains is that, because RNAi normally produces a 'knock-down' rather than a 'knock-out', it is also possible to identify interactions for essential genes.

C. elegans is a unique model animal in which genetic interactions can be identified *in vivo* in the context of a developing organism; the expression of any gene can be

systemically inhibited using long dsRNAs delivered by bacterial feeding (Timmons and Fire, 1998). Although it should be possible to identify genetic interactions between genes using RNAi in cell culture for mammalian or fly cells, *C. elegans* is currently the only model organism in which this approach can be used *in vivo* on a comprehensive scale. In *C. elegans*, RNAi screens can be performed in liquid culture in 96-well plates (Lehner et al., 2006c; van Haaften et al., 2004) using the bacterial feeding library (Kamath et al., 2003). This allows RNAi screens to be performed at sufficient throughput to be able to test tens of thousands of gene pairs for their ability to interact genetically *in vivo*.

Using high-throughput RNAi screens in *C. elegans*, we recently constructed the first systematic genetic interaction network for any animal (Lehner et al., 2006b). We focussed on genes that function in signalling pathways, and tested >65 000 pairs of genes for their ability to interact *in vivo* using both genetic mutant query strains and combinatorial RNAi. In total we identified 351 pairs of genes that when inactivated in combination produced a synthetic nonviable phenotype.

Mechanistic interpretation of genetic interaction networks

Synthetic lethal interactions from small-scale studies have normally been interpreted as providing supporting evidence for two gene products acting either in the same biochemical pathway or in two parallel pathways that can functionally compensate for each other (Hartman et al., 2001). In their original paper, Tong et al. noted that although ~27% of the genetic interactions that they identified link genes with similar Gene Ontology (GO) annotations, only ~1% of synthetic lethal interactions occur between genes whose products reside in the same protein complex (Tong et al., 2004). Rather, they demonstrated that genes encoding products that function in the same protein complex or pathway often have similar profiles of genetic interactions (i.e. genes from a single pathway tend to interact with the same genes, rather than with each other). Indeed the more genetic interactions two genes share, the more likely those two gene products are to interact physically (see below) (Tong et al., 2004).

By combining genetic interaction data with comprehensive protein–protein, protein–DNA and metabolic network data, Kelley and Ideker systematically compared the ability of ‘within-pathway’ models (also called ‘intra-pathway’ or ‘series’ models, Fig. 2A) or ‘between-pathway’ (‘inter-pathway’ or ‘parallel’) models (Fig. 2B) to explain systematically compiled genetic interaction data (Kelley and Ideker, 2005). Using a probabilistic model, they found that between-pathway models could explain three-and-a-half times as many interactions as within-pathway models. They were, however, unable to provide a mechanistic interpretation for ~60% of the observed genetic interactions in yeast. Indeed the extensive genetic interactions identified from both the SGA and dSLAM studies suggest that many complex compensatory relationships can occur between seemingly unrelated cellular pathways. For example Pan et al. observed extensive functional

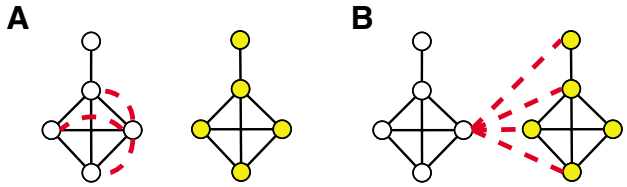

Fig. 2. Within- and between-pathway models for genetic interactions. Synthetic lethal interactions (broken lines) can occur both between two components of a single biochemical pathway (A), or between components of two parallel pathways that can functionally compensate for each other (B). Kelley and Ideker found that the combination of within- and between-pathway models could explain about 40% of synthetic lethal or sick interactions in yeast, with between-pathway models predominating. Examples of within-pathway interactions include interactions among components of the spliceosome, and interactions among components of the casein kinase 2 complex. Between-pathway interactions include extensive interactions between components of the Dynactin complex and components of the Prefoldin complex (Kelley and Ideker, 2005). For interactions between partial loss-of-function mutations, however, within-pathway models may predominate. Genes/proteins are shown as nodes, protein interactions as solid edges, and genetic interactions as broken edges.

compensation between loss of DNA damage response pathway genes and genes involved in mRNA transcription, mRNA processing and Golgi integrity (Pan et al., 2006). Therefore the mechanistic interpretation of genetic interactions remains an important area for future work.

In contrast to the situation in yeast, in the *C. elegans* genetic interaction network (Lehner et al., 2006b), within-pathway interactions appear to account for more interactions than between-pathway interactions: focussing on known components of signalling pathways, twice as many interactions are seen between components of the same pathway than between components of separate pathways (B.L., unpublished observation). The explanation for this probably lies in the difference in experimental approaches used: in yeast the interactions analysed are primarily between null alleles, whereas in *C. elegans* most interactions are between an hypomorphic allele and RNAi knockdown of a second gene. This makes intuitive sense: whereas the phenotypic consequences of a null mutation in a gene in a linear pathway cannot be further enhanced by a second mutation in that pathway, two partial loss-of-function mutations in a single pathway can be combined to inhibit that pathway completely (Fig. 2). Therefore with null alleles, many interactions probably represent interactions between the complete inactivation of two non-essential pathways that can functionally compensate for each other, whereas with hypomorphic alleles many of the interactions may represent interactions between the partial inactivation of two genes that act in the same essential pathway.

A further simple class of synthetic lethal interaction is seen between two duplicated genes that encode homologous proteins. Although many duplicated genes clearly do encode redundant functions that are maintained over considerable

evolutionary periods (Tischler et al., 2006), careful analysis of single gene mutant phenotypes (Wagner, 2005) and the SGA synthetic lethal data (Tong et al., 2004) suggests that gene duplications explain only a very small minority of synthetic lethal interactions [<2% of the interactions identified by SGA encode homologous proteins (Tong et al., 2004)].

Using genetic interactions to understand gene function

The tendency for genes acting either within- or between-pathways to interact genetically means that genetic interactions can be used to predict gene functions. As first noted by Tong et al., components of the same pathway tend to share similar synthetic lethal partners (Tong et al., 2004; Ye et al., 2005). Therefore the number of shared genetic interaction partners for two genes can be used to rank the probability of the two gene products physically interacting and sharing a biological function. The similarity of the interaction profiles of two genes can be mathematically defined by their 'congruence' (Ye et al., 2005). Congruence can be defined as the negative log of the hypergeometric *P*-value of the number of shared genetic interaction neighbours of two genes (Ye et al., 2005), which accounts for both the number of interactions each gene makes and the total size of the interaction network. Using yeast genetic interaction data, congruence performs better as a predictor of physical interactions and shared function than either direct genetic interactions or counting the number of interaction partners shared by two genes (Ye et al., 2005). Pan et al. were able to use congruence to define 16 functionally homogenous modules of DNA damage response genes (Pan et al., 2006) and were able to use the genetic interaction profiles of novel genes to predict their functions in the DNA damage response (Pan et al., 2006).

In the *C. elegans* genetic interaction network, because more interactions occur within a pathway than between pathways, knowledge of the direct interaction partners of a gene can be used to predict its function. For example, we systematically tested whether all of the genes that were found to interact with two or more known components of the EGF/Ras/MAPK pathway acted as general modulators of that pathway. For 9/16 of such cases tested, we found that the genes could indeed modulate EGF signalling in a precise developmental setting, suggesting that they do indeed act as general modulators of this pathway (Lehner et al., 2006b).

One way in which the accuracy of gene function prediction from genetic interaction networks can probably be improved is to identify both aggravating and alleviating interactions, and to quantify the strength of interactions. Indeed Schuldiner et al. found that including quantitative interaction data helped to define modules of genes that share precise molecular functions in the early secretory pathway (Schuldiner et al., 2005).

Predicting genetic interactions

Ideally we would like to construct a comprehensive genetic interaction network for humans. This network would then serve as a framework for directly predicting modifier genes in hereditary and somatic diseases, as well as providing a resource to predict human gene function on a genomic scale. However, identifying genetic interactions in multicellular organisms other than *C. elegans* is extremely laborious. Although it is possible to use RNAi to identify genetic interactions in cultured mammalian or fly cells (Wheeler et al., 2004), the inefficiency and expense of RNAi in mammalian cells (Pei and Tuschl, 2006), and the prevalence of off-target effects in both systems (Kulkarni et al., 2006; Lehner et al., 2004; Ma et al., 2006), has to date limited the application of this approach. Moreover it is not entirely clear how synthetic lethal interactions identified in single cells will relate to interactions in whole organisms.

An alternative to identifying genetic interactions using large-scale experimental approaches is to use computational methods to predict genetic interactions between genes. This is analogous to protein interaction networks, where methods that computationally predict protein interactions are now developed to the point that they are at least as accurate as most high-throughput experimental protocols or interactions derived from the literature (Jansen et al., 2003; Lee et al., 2004; Troyanskaya et al., 2003). Here I discuss three approaches that have been used for predicting genetic interactions: (i) using existing genetic interactions and the local network structure to predict new interactions, (ii) using the integration of other genomic datasets to predict genetic interactions, and (iii) using interactions from one species to predict interactions in a second species.

Predicting genetic interactions using network structure

One property of the yeast genetic interaction network is that two genes that share a genetic interaction with a common partner are likely to interact with each other (Tong et al., 2004). Tong et al. first exploited this 'small world' feature of genetic interaction networks to predict further interactions and found that in ~20% of cases the neighbours of a query gene could also interact with each other (compared to <1% of random gene pairs).

Predicting genetic interactions using other genomic datasets

Genes that share known functions are likely to have similar genetic interaction profiles. Genetic interactions can therefore be predicted using additional genomic datasets that link genes according to their functions. For example, genetic interactions for genes encoding proteins that physically interact have been successfully predicted (Kelley and Ideker, 2005; Ye et al., 2005). However a more powerful approach is to combine multiple different datasets that connect functionally related genes, and to use these to predict genetic interactions. Here I discuss the two approaches that have been applied to date: decision trees, and Bayesian integration.

Decision trees provide a method for classifying data into two or more classes (here 'interacting' and 'non-interacting') using multiple different evidence types. At each step in the tree a list of genes is divided into those that do or do not possess a particular characteristic. At the top of the tree the gene list is

first split using the characteristic that is most informative for predicting the property of interest (here the ability to predict a genetic interaction). Additional characteristics are then used to make additional subdivisions of the gene list until no additional characteristic is informative and a branch is terminated. An advantage of decision trees over 'black box' methods such as neural networks and support vector machines is that they explicitly reveal the characteristics used to classify the data. They also do not assume independence between predictive evidence types, which allows multiple related datasets to be used even if they contain correlations with each other. Wong et al. used decision trees to integrate protein localisation, mRNA expression, physical interaction, known function and network topology data in order to predict synthetic lethal or sick interactions between yeast genes (Wong et al., 2004). Using cross-validation tests, they found that decision trees could reliably predict genetic interactions between yeast genes. They also tested the predictions for eight new SGA screens not seen in training; 49/318 predictions were verified, compared to 2/318 expected by chance (Wong et al., 2004). The top predictor in the decision tree was the previously noted network property that genes that share genetic interaction partners are also likely to interact genetically. 'Between-pathway' models were also found to be useful predictors of interactions. However, individually omitting other kinds of functional data alone had little effect on the quality of predictions (Wong et al., 2004).

An alternative method for integrating genomic datasets to predict genetic interactions is Bayesian integration. In this approach predictions made using different data types are weighted according to the ability of each data type to predict known genetic interactions, as opposed to genes that are known not to interact (Jansen et al., 2003; Lee et al., 2004; Troyanskaya et al., 2003). Zhong and Sternberg recently used this approach to predict genetic interactions for ~10% of *C. elegans* genes, using information on anatomical expression patterns, phenotypes, functional annotations, microarray coexpression and protein interactions to predict genetic interactions (Zhong and Sternberg, 2006). The authors used their network to identify twelve subtle modifiers of mutations in the *let-60* Ras gene and two novel suppressors of mutations in the *itr-1* gene. Using a modified Bayesian integration, we have extended this approach to construct a network of >100 000 interactions that covers >60% of *C. elegans* genes, and have used this network to identify new suppressors of mutations in the Retinoblastoma pathway and cross-talk between the Dystrophin complex and the EGF/Ras/MAPK pathway (I. Lee, B.L., C. Crombie, A. Fraser and E. M. Marcotte, unpublished data).

Predicting genetic interactions using orthology relationships

A final approach for predicting genetic interactions is to use genetic interactions identified in one species to predict genetic interactions in a second species. The evolutionary conservation of protein–protein interactions between species (Matthews et al., 2001), means that human protein–protein interactions can be successfully predicted by using data from model organisms (Lehner and Fraser, 2004). To test whether genetic interactions can also be successfully transferred between species, we have tested whether the orthologs of genes that are synthetically lethal in *S. cerevisiae* are also synthetically lethal in *C. elegans*. In total we have tested >1000 predicted interactions, but have found that <1% are conserved (J. Tischler, B.L. and A. Fraser, unpublished data). This is in contrast to mutations in single genes; >60% of the orthologs of genes that are essential in *S. cerevisiae* are also essential in *C. elegans* (Kamath et al., 2003). Therefore at least for synthetic lethal interactions, interactions are probably not directly conserved between unicellular and multicellular organisms. This may reflect the presence of additional compensatory pathways in higher eukaryotes, or alternatively suggests that there is little evolutionary selection for 'higher order' interactions between mutations in pairs of genes. This does not mean, however, that synthetic lethal interactions between yeast genes are uninformative for predicting human gene function – clearly the synthetic lethal profiles of yeast genes are highly informative for predicting the molecular functions of orthologous human genes (Pan et al., 2006), and this information could be used to predict genetic interactions in humans. In future work it will be important to test whether genetic interactions can be successfully transferred between more related species (for example between *C. elegans* and humans). Anecdotal examples from the literature [for example the conservation of genetic interactions between components of the EGF/Ras/MAPK pathway between worms and flies (Sundaram, 2005)], suggest that this should be more successful. In addition it is possible to envisage more sophisticated approaches, whereby interactions between pathways, rather than genes, are predicted.

The implications of genetic interaction networks for human disease: hubs, buffers and new paradigms for human disease

What are the implications of model organism genetic interaction networks for human genetic diseases? The most immediate observation is that genetic interaction networks are very dense – there are many more ways to produce a phenotype by combining mutations in two genes than by mutating a single gene (Tong et al., 2004). This suggests that most hereditary diseases are also likely to result primarily from interactions between mutations in multiple genes rather than from mutations in single genes (Badano and Katsanis, 2002). Indeed the prevalence of synthetic interactions between genes may explain the relative lack of success in identifying the causal mutations in most common hereditary diseases – each disease may result from many different combinations of mutations in many different genes, each of which has only a minor or zero effect on the disease alone. This paints a rather pessimistic picture for the practicalities of genetically dissecting and treating complex genetic diseases.

However the structure of model organism genetic interaction networks also has a positive implication for human disease

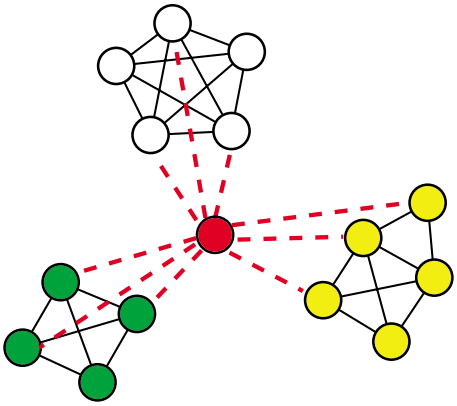

Fig. 3. Genetic hubs and genetic disease in humans. Genetic hubs are genes that when inactivated can enhance the phenotypic consequence of mutations in many different genes. Often hub genes can enhance the consequences of mutations in genes acting in diverse functionally unrelated pathways. Examples include a set of chromatin-modifying genes in *C. elegans* (the genes *mys-1*, *trr-1*, *dpy-22*, *hmg-1.2*, *din-1*, and *egl-27*) (Lehner et al., 2006b), the Prefoldin complex in *S. cerevisiae* (Tong et al., 2004), and the gene *hsp90* in yeast (Zhao et al., 2005), flies (Rutherford and Lindquist, 1998) and plants (Queitsch et al., 2002). Here the red node represents a hub gene, and the remaining nodes are coloured according to their function. Protein–protein interactions are shown as solid lines and genetic interactions as broken lines.

genetics – the existence of highly connected 'hub' genes in the networks suggests that there will be common modifiers of genetic mutations in humans. For example, in the *C. elegans* genetic interaction network we identified a class of genes that interact genetically with many diverse genes. Inactivation of each of these 'hub' genes can enhance the loss-of-function phenotype resulting from mutations in many different genes with diverse molecular functions (Fig. 3). Indeed, loss of a hub gene can enhance many different phenotypes, depending upon the other gene that is mutated in combination. In this way they can be thought of as 'buffering' an organism from the phenotypic consequences of mutations. Remarkably all of the most connected hub genes that we identified function in chromatin-modifying complexes (Lehner et al., 2006b).

There are probably many more hub genes in addition to these chromatin-modifying genes. For example, in the SGA yeast genetic interaction network, four of the top five most connected genes encode components of the prefoldin chaperone complex (Tong et al., 2004). Moreover the chaperone *hsp90* can also be classed as a hub gene, because of its ability to enhance the phenotypic consequences of mutations in multiple genes when inactivated in flies, plants and yeast (Queitsch et al., 2002; Rutherford and Lindquist, 1998; Zhao et al., 2005). Thus at least two classes of genes can function as genetic hubs: chromatin-modifiers and chaperones. Interestingly *hsp90* may bridge these two functional classes; although it is well known as a chaperone, it may also affect phenotypic variation *via* its effects on chromatin structure (Sollars et al., 2003).

Although there are likely many other hub genes that remain to be identified, their implication for human disease is clear: there will probably be human genes that act as modifier genes in many mechanistically unrelated diseases. Indeed the concept of hub genes suggests a new paradigm for genetic disease in humans (Lehner et al., 2006b). In this paradigm there are two classes of human disease gene: the first class consists of 'specifier' genes that define the particular disease, and the second class consists of 'modifier' or 'hub' genes that serve to enhance the strength of the disease resulting from a mutation in a specifier gene. There is good evidence to suggest that hub genes identified in one organism also function as hubs in other species (Lehner et al., 2006b; Queitsch et al., 2002; Zhao et al., 2005), and so the particular genes identified as hubs in model organisms may also function as hub genes in humans.

Work in my lab is funded by the EMBL-CRG Systems Biology Program, which is supported by a grant from the Spanish Ministry of Science and Education (Ministerio de Educación y Ciencia, MEC).

References

Badano, J. L. and Katsanis, N. (2002). Beyond Mendel: an evolving view of human genetic disease transmission. *Nat. Rev. Genet.* **3**, 779-789.

Baugh, L. R., Wen, J. C., Hill, A. A., Slonim, D. K., Brown, E. L. and Hunter, C. P. (2005). Synthetic lethal analysis of *Caenorhabditis elegans* posterior embryonic patterning genes identifies conserved genetic interactions. *Genome Biol.* **6**, R45.

Collins, S. R., Schuldiner, M., Krogan, N. J. and Weissman, J. S. (2006). A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* **7**, R63.

Davierwala, A. P., Haynes, J., Li, Z., Brost, R. L., Robinson, M. D., Yu, L., Mnaimneh, S., Ding, H., Zhu, H., Chen, Y. et al. (2005). The synthetic genetic interaction spectrum of essential genes. *Nat. Genet.* **37**, 1147-1152.

Drees, B. L., Thorsson, V., Carter, G. W., Rives, A. W., Raymond, M. Z., Avila-Campillo, I., Shannon, P. and Galitski, T. (2005). Derivation of genetic interaction networks from quantitative phenotype data. *Genome Biol.* **6**, R38.

Eason, R. G., Pourmand, N., Tongprasit, W., Herman, Z. S., Anthony, K., Jejelowo, O., Davis, R. W. and Stolc, V. (2004). Characterization of synthetic DNA bar codes in *Saccharomyces cerevisiae* gene-deletion strains. *Proc. Natl. Acad. Sci. USA* **101**, 11046-11051.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B. et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-391.

Hartman, J. L., 4th and Tippery, N. P. (2004). Systematic quantification of gene interactions by phenotypic array analysis. *Genome Biol.* **5**, R49.

Hartman, J. L., 4th, Garvik, B. and Hartwell, L. (2001). Principles for the buffering of genetic variation. *Science* **291**, 1001-1004.

Holway, A. H., Hung, C. and Michael, W. M. (2005). Systematic, RNA-interference-mediated identification of mus-101 modifier genes in *Caenorhabditis elegans*. *Genetics* **169**, 1451-1460.

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., Emili, A., Snyder, M., Greenblatt, J. F. and Gerstein, M. (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* **302**, 449-453.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M. et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231-237.

Kelley, R. and Ideker, T. (2005). Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.* **23**, 561-566.

Kulkarni, M. M., Booker, M., Silver, S. J., Friedman, A., Hong, P., Perrimon, N. and Mathey-Prevot, B. (2006). Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nat. Methods* **3**, 833-838.

Lee, I., Date, S. V., Adai, A. T. and Marcotte, E. M. (2004). A probabilistic functional network of yeast genes. *Science* **306**, 1555-1558.

Lehner, B. and Fraser, A. G. (2004). A first-draft human protein-interaction map. *Genome Biol.* **5**, R63.

Lehner, B., Fraser, A. G. and Sanderson, C. M. (2004). Technique review: how to use RNA interference. *Brief. Funct. Genomic. Proteomic.* **3**, 68-83.

Lehner, B., Calixto, A., Crombie, C., Tischler, J., Fortunato, A., Chalfie, M. and Fraser, A. G. (2006a). Loss of LIN-35, the *Caenorhabditis elegans* of the tumor suppressor p105Rb, results in enhanced RNA interference. *Genome Biol.* **7**, R4.

Lehner, B., Crombie, C., Tischler, J., Fortunato, A. and Fraser, A. G. (2006b). Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat. Genet.* **38**, 896-903.

Lehner, B., Tischler, J. and Fraser, A. G. (2006c). RNAi screens in *C. elegans* in a 96-well liquid format and their application to the systematic identification of genetic interactions. *Nat. Protoc.* **1**, 1617-1620.

Ma, Y., Creanga, A., Lum, L. and Beachy, P. A. (2006). Prevalence of off-target effects in Drosophila RNA interference screens. *Nature* **443**, 359-363.

Matthews, L. R., Vaglio, P., Reboul, J., Ge, H., Davis, B. P., Garrels, J., Vincent, S. and Vidal, M. (2001). Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or 'interologs'. *Genome Res.* **11**, 2120-2126.

Ooi, S. L., Shoemaker, D. D. and Boeke, J. D. (2003). DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. *Nat. Genet.* **35**, 277-286.

Ooi, S. L., Pan, X., Peyser, B. D., Ye, P., Meluh, P. B., Yuan, D. S., Irizarry, R. A., Bader, J. S., Spencer, F. A. and Boeke, J. D. (2006). Global synthetic-lethality analysis and yeast functional profiling. *Trends Genet.* **22**, 56-63.

Pan, X., Yuan, D. S., Xiang, D., Wang, X., Sookhai-Mahadeo, S., Bader, J. S., Hieter, P., Spencer, F. and Boeke, J. D. (2004). A robust toolkit for functional profiling of the yeast genome. *Mol. Cell* **16**, 487-496.

Pan, X., Ye, P., Yuan, D. S., Wang, X., Bader, J. S. and Boeke, J. D. (2006). A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell* **124**, 1069-1081.

Pei, Y. and Tuschl, T. (2006). On the art of identifying effective and specific siRNAs. *Nat. Methods* **3**, 670-676.

Pierce, S. E., Fung, E. L., Jaramillo, D. F., Chu, A. M., Davis, R. W., Nislow, C. and Giaever, G. (2006). A unique and universal molecular barcode array. *Nat. Methods* **3**, 601-603.

Queitsch, C., Sangster, T. A. and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618-624.

Rutherford, S. L. and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336-342.

Schuldiner, M., Collins, S. R., Thompson, N. J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J. F. et al. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507-519.

Sollars, V., Lu, X., Xiao, L., Wang, X., Garfinkel, M. D. and Ruden, D. M. (2003). Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nat. Genet.* **33**, 70-74.

Sundaram, M. V. (2005). RTK/Ras/MAP kinase signaling. In *WormBook* (ed. The *C. elegans* Research Community), doi/10.1895/wormbook.1.80.1, http://www.wormbook.org.

Suzuki, Y. and Han, M. (2006). Genetic redundancy masks diverse functions of the tumor suppressor gene PTEN during *C. elegans* development. *Genes Dev.* **20**, 423-428.

Timmons, L. and Fire, A. (1998). Specific interference by ingested dsRNA. *Nature* **395**, 854.

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

Tong, A. H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C. W., Bussey, H. et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364-2368.

Tong, A. H., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G. F., Brost, R. L., Chang, M. et al. (2004). Global mapping of the yeast genetic interaction network. *Science* **303**, 808-813.

Troyanskaya, O. G., Dolinski, K., Owen, A. B., Altman, R. B. and Botstein, D. (2003). A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*). *Proc. Natl. Acad. Sci. USA* **100**, 8348-8353.

van Haaften, G., Vastenhouw, N. L., Nollen, E. A., Plasterk, R. H. and Tijsterman, M. (2004). Gene interactions in the DNA damage-response pathway identified by genome-wide RNA-interference analysis of synthetic lethality. *Proc. Natl. Acad. Sci. USA* **101**, 12992-12996.

Wagner, A. (2005). Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays* **27**, 176-188.

Wheeler, D. B., Bailey, S. N., Guertin, D. A., Carpenter, A. E., Higgins, C. O. and Sabatini, D. M. (2004). RNAi living-cell microarrays for loss-of-function screens in *Drosophila melanogaster* cells. *Nat. Methods* **1**, 127-132.

Wong, S. L., Zhang, L. V., Tong, A. H., Li, Z., Goldberg, D. S., King, O. D., Lesage, G., Vidal, M., Andrews, B., Bussey, H. et al. (2004). Combining biological networks to predict genetic interactions. *Proc. Natl. Acad. Sci. USA* **101**, 15682-15687.

Ye, P., Peyser, B. D., Pan, X., Boeke, J. D., Spencer, F. A. and Bader, J. S. (2005). Gene function prediction from congruent synthetic lethal interactions in yeast. *Mol. Syst. Biol.* **1**, 2005 0026.

Yuan, D. S., Pan, X., Ooi, S. L., Peyser, B. D., Spencer, F. A., Irizarry, R. A. and Boeke, J. D. (2005). Improved microarray methods for profiling the Yeast Knockout strain collection. *Nucleic Acids Res.* **33**, e103.

Zhao, R., Davey, M., Hsu, Y. C., Kaplanek, P., Tong, A., Parsons, A. B., Krogan, N., Cagney, G., Mai, D., Greenblatt, J. et al. (2005). Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* **120**, 715-727.

Zhong, W. and Sternberg, P. W. (2006). Genome-wide prediction of *C. elegans* genetic interactions. *Science* **311**, 1481-1484.

The Journal of Experimental Biology 210, 1492-1496
Published by The Company of Biologists 2007
doi:10.1242/jeb.02783

Glossary of terms

This section is designed to help readers adapt to the complex terminology associated with contemporary molecular genetics, genomics and systems biology. Fuller descriptions of these terms are available at http://www.wikipedia.org/

Ab initio prediction	methods used to predict the potential genes encoded in the genome, which are trained on datasets made of known genes, and used computationally to predict coding regions out of genome without the aid of cDNA sequence. Although their performance is improving, these algorithms perform very poorly on non-protein coding genes.
Annotation	as applied to proteins, DNA sequences or genes. The storage of data describing these entities (protein/gene identities, DNA motifs, gene ontology categorisation, etc.) within a biological database. Active projects include FlyBase and WormBase. See Gene ontology.
Assembly	the process of aligning sequenced fragments of DNA into their correct positions within the chromosome or transcript.
cDNA	complementary DNA. This is DNA synthesised from a mature mRNA template by the enzyme reverse transcriptase. cDNA is frequently used as an early part of gene cloning procedures, since it is more robust and less subject to degradation than the mRNA itself.
ChIP	chromatin immunoprecipitation assay used to determine which segments of genomic DNA are bound to chromatin proteins, mainly including transcription factors.
Chip	see Microarray.
ChIP-on-chip	use of a DNA microarray to analyse the DNA generated from chromatin immunoprecipitation experiments (see ChIP).
cis-acting	a molecule is described as *cis*-acting when it affects other genes that are physically adjacent, on the same chromosome, or are genetically linked or in close proximity (for mRNA expression, typically a promoter).
Collision-induced dissociation	a mechanism by which molecules (e.g. proteins) are fragmented to form molecular ions in the gas phase. These fragments are then analysed within a mass spectrometer to provide mass determination.
Connectivity	a term from graph theory, which indicates the number of connections between nodes or vertices in a network. Greater connectedness between nodes is generally used as a measure of robustness of a network.
CpG islands	regions that show high density of ‘C followed by G’ dinucleotides and are generally associated with promoter elements; in particular, stretches of DNA of at least 200 bp with a C–G content of 50% and an observed CpG/expected CpG in excess of 0.6. The cytosine residues can be methylated, generally to repress transcription, while demethylated CpGs are a hallmark of transcription. CpG dinucleotides are under-represented outside regulatory regions, such as promoters, because methylated C mutates into T by deamination.
Edge	as in networks. Connects two nodes (or vertices) within a system. These concepts arise from graph theory.
Enhancer	a short segment of genomic DNA that may be located remotely and that, on binding particular proteins (*trans*-acting factors), increases the rate of transcription of a specific gene or gene cluster.
Epistasis	a phenomenon when the properties of one gene are modified by one or more genes at other loci. Otherwise known as a genetic interaction, but epistasis refers to the statistical properties of the phenomenon.

eQTL	the combination of conventional QTL analysis with gene expression profiling, typically using microarrays. eQTLs describe regulatory elements controlling the expression of genes involved in specific traits.
EST	expressed sequence tag. A short DNA sequence determined for a cloned cDNA representing portions of an expressed gene. The sequence is generally several hundred base pairs from one or both ends of the cloned insert.
Exaptation	a biological adaptation where the current function is not that which was originally evolved. Thus, the defining (derived) function might replace or persist with the earlier, evolved adaptation.
Exon	any region of DNA that is transcribed to the final (spliced) mRNA molecule. Exons interleave with segments of non-coding DNA (introns) that are removed (spliced out) during processing after transcription.
Gene forests	genomic regions for which RNA transcripts, produced from either DNA strand, have been identified without gaps (non-transcribed genomic regions). Conversely, regions in which no transcripts have ever been detected are called ‘gene deserts’.
Gene interaction network	a network of functional interactions between genes. Functional interactions can be inferred from many different data types, including protein–protein interactions, genetic interactions, co-expression relationships, the co-inheritance of genes across genomes and the arrangement of genes in bacterial genomes. The interactions can be represented using network diagrams, with lines connecting the interacting elements, and can be modelled using differential equations.
Gene ontology (GO)	an ontology is a controlled vocabulary of terms that have logical relationships with each other and that are amenable to computerised manipulation. The Gene Ontology project has devised terms in three domains: biological process, molecular function and cell compartment. Each gene or DNA sequence can be associated with these annotation terms from each domain, and this enables analysis of microarray data on groups of genes based on descriptive terms so provided. See http://www.geneontology.org
Gene set enrichment analysis	a computational method that determines whether a defined set of genes, usually based on their common involvement in a biological process, shows statistically significant differences in transcript expression between two biological states.
Gene silencing	the switching-off of a gene by an epigenetic mechanism at the transcriptional or post-transcriptional levels. Includes the mechanism of RNAi.
Genetic interaction (network)	a genetic interaction between two genes occurs when the phenotypic consequences of a mutation in one gene are modified by the mutational status at a second locus. Genetic interactions can be aggravating (enhancing) or alleviating (suppressing). To date, most high-throughput studies have focussed on systematically identifying synthetic lethal or sick (aggravating) interactions, which can then be visualised as a network of functional interactions (edges) between genes (nodes).
Genome	a portmanteau of gene and chromosome, the entire hereditary information for an organism that is embedded in the DNA (or, for some viruses, in RNA). Includes protein-coding and non-coding sequences.
Heritability	phenotypic variation within a population is attributable to the genetic variation between individuals and to environmental factors. Heritability is the proportion due to genetic variation usually expressed as a percentage.
Heterologous hybridization	the use of a cDNA or oligonucleotide microarray of probes designed for one species with target cRNA/cDNAs from a different species.
Homeotic	the transformation of one body part to another due to mutation of specific developmentally related genes, notably the *Hox* genes in animals and *MADS-box* genes in plants.
Hub	as in networks. A node with high connectivity, and thus which interacts with many other nodes in the network. A hub protein interacts with many other proteins in a cell.

Hybridisation	the process of joining (annealing) two complementary single-stranded DNAs into a single double-stranded molecule. In microarray analysis, the target RNA/DNA from the subject under investigation is denatured and hybridised to probes that are immobilised on a solid phase (i.e. glass microscope slide).
Hypomorph	in genetics, a loss-of-function mutation in a gene, but which shows only a partial reduction in the activity it influences rather than a complete loss (cf. hypermorph, antimorph, neomorph, etc).
Imprinting	a phenomenon where two inherited copies of a gene are regulated in opposite ways, one being expressed and the other being repressed.
Indel	insertion and deletion of DNA, referring to two types of genetic mutation. To be distinguished from a 'point mutation', which refers to the substitution of a single base.
Interactome	a more or less comprehensive set of interactions between elements within cells. Usually applied to genes or proteins as defined by transcriptomic, proteomic or protein–protein interaction data.
Intron	see Exon.
KEGG	The Kyoto Encyclopedia of Genes and Genomes is a database of metabolic and other pathways collected from a variety of organisms. See http://www.genome.jp/kegg
Metabolomics	the systematic qualitative and quantitative analysis of small chemical metabolite profiles. The metabolome represents the collection of metabolites within a biological sample.
Metagenomics	the application of genomic techniques to characterise complex communities of microbial organisms obtained directly from environmental samples. Typically, genomic tags are sequence characterised as markers of each species to inform on the range and abundance of species in the community.
Microarray	an arrayed set of probes for detecting molecularly specific analytes or targets. Typically, the probes are composed of DNA segments that are immobilised onto the solid surface, each of which can hybridise with a specific DNA present in the target preparation. DNA microarrays are used for profiling of gene transcripts.
Model species	a species used to study particular biological phenomena, the outcome offering insights into the workings of other species. Usually, the selection is based on experimental tractability, particularly ease of genetic manipulation. For the geneticist, it is an organism with inbred lines where sibs will be >98% identical (i.e. *Drosophila*, *Caenorhabditis elegans* and mice). For genomic science, it refers to a species for which the genomic DNA has been sequenced.
miRNA	a category of novel, very short, non-coding RNAs, generated by the cleavage of larger precursors (pri-miRNA). These short RNAs are included in the RNA-induced silencing complex (RISC) and pair to the 3′ ends of target RNA, blocking its translation into proteins (in animals) or promoting RNA cleavage and degradation (in plants).
mRNA	a protein-coding mRNA containing a protein-coding region (CDS), preceded by a 5′ and followed by a 3′ untranslated region (5′ UTR and 3′ UTR). The UTRs contain regulatory elements. A full-length cDNA contains the complete sequence of the original mRNA, including both UTRs. However, it is often difficult to assign the starting–termination positions for protein synthesis unambiguously. A cDNA containing the entire CDS is often considered acceptable for bioinformatic and experimental studies requiring full-length cDNAs.
ncRNA	non-coding RNA is any RNA molecule with no obvious protein-coding potential for at least 80 or 100 amino acids, as determined by scanning full-length cDNA sequences. It includes ribosomal (rRNA) and transfer RNAs (tRNA) and is now known to include various sub-classes of RNA, including snoRNA, siRNA and piRNA. Just like the coding mRNAs, a large proportion of ncRNAs are transcribed by RNA polymerase II and are large transcripts. A description of the many forms of ncRNA can be found at http://en.wikipedia.org/wiki/Non-coding_RNA.

Node	as in networks. Objects linked by edges to create a network.
PCR	polymerase chain reaction. A molecular biology technique for replicating DNA *in vitro*. The DNA is thus amplified, sometimes from very small amounts. PCR can be adapted to perform a wide variety of genetic manipulations.
piRNA	Piwi-interacting RNA. A class of RNA molecules (29–30 nt long) that complex with Piwi proteins (a class of the Argonaute family of proteins) and are involved in transcriptional gene silencing.
PMF	peptide mass fingerprinting. An analytical technique for protein identification in which a protein is fragmented using proteases. The resulting peptides are analysed by mass spectrometry and these masses compared against a database of predicted or measured masses to generate a protein identity.
Polyadenylation	the covalent addition of multiple A bases to the 3′ tail of an mRNA molecule. This occurs during the processing of transcripts to form the mature, spliced molecule and is important for regulation of turnover, trafficking and translation.
Post-source decay	in mass spectrometry. The fragmentation of precursor molecular ions as they accelerate away from the ionisation source of the mass spectrometer. All precursor ions leaving the ion source have approximately the same kinetic energy, but fragmentation results in smaller product ions that can be distinguished from precursor ions using a ‘reflectron’ by virtue of their lower kinetic energies.
Post-translational modification	the chemical modification of a protein after synthesis through translation. Some modifications, notably phosphorylation, affect the properties of the protein, offering a means of regulating function.
Principal component analysis (PCA)	a technique for simplifying complex, multi-dimensional datasets to a reduced number of dimensions, the principal components. This procedure retains those characteristics of the data that relate to its variance.
Promoter	a regulatory DNA sequence, generally lying upstream of an expressed gene, which in concert with other often distant regulatory elements directs the transcription of a given gene.
Proteome	the entire protein complement of an organism, tissue or cell culture at a given time.
Quantitative trait	inheritance of a phenotypic property or characteristic that varies continuously between extreme states and can be attributed to interactions between multiple genes and their environment.
qPCR	quantitative real-time PCR, sometimes called real-time PCR. A more quantitative form of RT-PCR in which the quantity of amplified product is estimated after each round of amplification.
QTL	quantitative trait loci. A region of DNA that contains those genes contributing to the trait under study.
RISC	RNA-induced silencing complex. A protein complex that mediates the double-stranded RNA-induced destruction of homologous mRNA.
RNAi	RNA interference or RNA-mediated interference. The process by which double-stranded RNA triggers the destruction of homologous mRNA in eukaryotic cells by the RISC.
RT-PCR	reverse transcription–polymerase chain reaction. A technique for amplifying a defined piece of RNA that has been converted to its complementary DNA form by the enzyme reverse transcriptase. See qPCR.
siRNA	small interfering RNA, or silencing RNA. A class of short (20–25 nt), double-stranded RNA molecules. It is involved in the RNA interference pathway, which alters RNA stability and thus affects RNA concentration and thereby suppresses the normal expression of specific genes. Widely used in biomedical research to ablate specific genes.

snoRNA	small nucleolar RNA. A sub-class of RNA molecules involved in guiding chemical modification of ribosomal RNA and other RNA genes as part of the regulation of gene expression.
SNP	single nucleotide polymorphism. A single base-pair mutation at a specific locus, usually consisting of two alleles. Because SNPs are conserved over evolution, they are frequently used in QTL analysis and in association studies in place of microsatellites, and in genetic fingerprinting analyses.
SSH	suppressive subtractive hybridisation. A powerful protocol for enriching cDNA libraries for genes that differ in representation between two or more conditions. It combines normalisation and subtraction in a single procedure and allows the detection of low-abundance, differentially expressed transcripts, such as those involved in signalling and signal transduction.
Structural RNAs	a class of non-coding RNA, long known to have a structural role (for instance, the ribosomal RNAs), transcribed by RNA polymerase I or III.
Systems biology	treatment of biological entities as systems composed of defined elements interacting in defined ways to enable the observed function and behaviour of that system. The properties of the systems are embedded in a quantitative model that guides further tests of systems behaviour.
TATA-boxes	sequences in promoter regions constituted by TATAAA, or similar variants, which were considered the hallmark of Promoters. Recent data show that they are present only in the minority of promoters, where they direct transcription at a single well-defined location some 30 bp downstream of this element.
trans-acting	a factor or gene that acts on another unlinked gene, a gene on a separate chromosome or genetically unlinked usually through some diffusible protein product (for mRNA expression, typically a transcription factor).
Transcript	an RNA product produced by the action of RNA polymerase reading the sequence of bases in the genomic DNA. Originally limited to protein-coding sequences with flanking UTRs but now known to include large numbers of products that do not code for a protein product.
Transcriptome	the full set of mRNA molecules (transcripts) produced by the system under observation. Whilst the genome is fixed for a given organism, the transcriptome varies with context (i.e. tissue source, ontogeny, external conditions or experimental treatment).
Transgene	a gene or genetic material that has been transferred between species or between organisms using one of several genetic engineering techniques.
Transinduction	generation of transcripts from intergenic regions. At least some such products do not relate to a definable promoter or transcriptional start site.
Transposon	sequences of DNA able to move to new positions within the genome of a single cell. This event might cause mutation at the site of insertion. Also called ‘mobile genetic elements’ or ‘jumping genes’.
Transvection	an epigenetic phenomenon arising from the interaction between one allele and the corresponding allele on the homologous chromosome, leading to gene regulation.
TUs	transcriptional units. Used to group all of the overlapping RNA transcripts that are transcribed from the same genomic strand and share exonic sequences.
UTR	untranslated region. Regions of the mRNA that lie at either the 3′ or 5′ flanking ends of the molecule (i.e. 3′ UTR and 5′ UTR). They bracket the protein-coding region and contain signals and binding sites that are important for the regulation of both protein translation and RNA degradation.