Integrative approaches for finding modular structure in biological networks

Koyel Mitra[1], Anne-Ruxandra Carvunis[1]*, Sanath Kumar Ramesh[2] and Trey Ideker[1,2,3]*

Abstract | A central goal of systems biology is to elucidate the structural and functional architecture of the cell. To this end, large and complex networks of molecular interactions are being rapidly generated for humans and model organisms. A recent focus of bioinformatics research has been to integrate these networks with each other and with diverse molecular profiles to identify sets of molecules and interactions that participate in a common biological function — that is, 'modules'. Here, we classify such integrative approaches into four broad categories, describe their bioinformatic principles and review their applications.

Epistasis
The phenomenon whereby the function of one gene affects the phenotype (for example, growth) of another gene in a non-additive manner.

Synthetic lethality
An extreme case of negative genetic epistasis in which the mutation of two genes in combination, but not individually, causes a lethal phenotype.

[1]Department of Medicine, University of California San Diego, La Jolla, California 92093, USA.
[2]Department of Computer Science and Engineering, University of California San Diego, La Jolla, California 92093, USA.
[3]Department of Bioengineering, University of California San Diego, La Jolla, California 92093, USA.
**These authors contributed equally to this work.*
Correspondence to K.M.
e-mail: kmitra@ucsd.edu
doi:10.1038/nrg3552

Cellular organization is thought to be fundamentally modular[1,2]. At the molecular level, modules have been variously described as groups of genes, gene products or metabolites that are functionally coordinated, physically interacting and/or co-regulated[1–7]. A pioneering perspective[1] on modular cell biology described a module as a distinct group of interacting molecules driving a common biological process — for example, the ribosome is a module that synthesizes proteins. Modules, in essence, are functional building blocks of the cell[1–7].

In an effort to develop a complete map of biological modules underlying cellular architecture and function, large networks of intermolecular interactions are being measured systematically for humans and many model species[8–16]. Such networks include physical associations underlying protein–protein, protein–DNA or metabolic pathways, as well as functional associations, including epistasis and synthetic lethality relationships between genes, correlated expression between genes, or correlated biochemical activities among other types of molecules (Supplementary information S1 (table)). Numerous approaches have been developed to mine such networks for identifying biological modules, including methods for clustering interactions and those based on topological features of the network such as degree and betweenness centrality (as reviewed in REFS 5–7). These approaches are based on the premise that modular structures such as protein complexes, signalling cascades or transcriptional regulatory circuits display characteristic patterns of interaction[5–7]. They have been extremely powerful for elucidating molecular machineries underlying physiological and disease phenotypes[5–7,17–19].

Nonetheless, many challenges confound the interpretation of biological networks and their embedded modular structures. A first challenge relates to the sheer complexity of the problem at hand: it is not yet clear how to transform data for thousands of molecular interactions into functionally coherent models of cellular machinery. Second, technological biases in high-throughput approaches[20–22] can compromise signal accuracy. Experimental artefacts, variability in coverage across data sets, sampling bias towards well-studied processes, limitations in screening power and inherent sensitivities in various assays can yield false positives and false negatives in interaction data[23–26]. Third, individual high-throughput experiments measuring a subset or type of interactions (for example, protein–protein or protein–DNA interactions) simply cannot expose the full interaction landscape of a cell. Finally, as molecular networks are commonly assembled in single, static experimental conditions, they overlook the inherently dynamic nature of molecular interactions, which can be extensively rewired during physiological or environmental shifts[10,27,28]. Hence, current network models reveal only partial and static snapshots of the cell.

A key strategy to address these challenges is data integration. In recent years, a rich collection of integrative methods has emerged for the identification of network modules of high quality and broad coverage, and of context-specific dynamics. Here, we review

these integrative approaches, highlighting their logical underpinnings and biological applications. We classify integrative module discovery methods into four broad categories: identification of 'active modules' through the integration of networks and molecular profiles, identification of 'conserved modules' across multiple species, identification of 'differential modules' across different conditions and identification of 'composite modules' through the integration of different interaction types. Together, these four categories encompass a wide range of network integration strategies and available data types. An illustrative poster[29] titled 'Integrative Systems Biology' was previously published and is recommended as an accompanying guide.

Identification of active modules

One of the most successful integrative approaches has been to overlay networks with molecular profiles to identify 'active modules'. Molecular profiles of transcriptomic, genomic, proteomic, epigenomic and other cellular information are rapidly populating public databases (Supplementary information S1 (table)). As these profiles capture dynamic and process-specific information that is correlated with cellular or disease states, they naturally complement interaction data, which are primarily derived under a single experimental condition. Computational integration of network and 'omics' profiles has thus become a popular strategy for extracting context-dependent active modules, which mark regions of the network showing striking changes in molecular activity (for example, transcriptomic expression) or phenotypic signatures (for example, mutational abundance) that are associated with a given cellular response[4,30–38] (FIG. 1). These regions have alternatively been described as network hotspots[39,40] or responsive subnetworks[41–43].

Many computational techniques have been developed that automate the large-scale identification of active modules in an unbiased manner. Several of these methods have been packaged as publicly available application tools (TABLE 1). These methods generally fall into three classes, as described in the following subsections. Given the rapid emergence of integrative methodologies, some effort has been made to compare their accuracy (precision), sensitivity (recall) or computational efficiency within individual method classes[44–46]. However, unbiased comparisons across different classes of methods using uniform data metrics will need to be undertaken comprehensively[47].

Significant-area-search methods. The first class of methods, themed SigArSearch (significant area search)[31,33,48] was previously reviewed[43]. Many of these methods[33,41,44,48–56] descend from an early formulation, jActiveModules[48], (also implemented as an application tool through the network analysis and visualization platform, Cytoscape[57] (TABLE 1)), which was the first to frame the active modules search task as an optimization problem. SigArSearch methods invoke three common procedural steps for module discovery (FIG. 1). First, network nodes (molecules) and/or edges (interactions) are annotated with scores that quantify molecular activity, which is measured using molecular profiles such as gene expression levels (the most common data choice in such applications). Next, a scoring function is formulated to compute an aggregate score for each subnetwork that reflects the overall activity of member nodes and interactions. Subsequently, a search strategy is devised to identify subnetworks with high scores, which mark active modules.

Scoring and searching for active modules present a range of computational considerations and implementations[43]. Different scoring functions have assumed scores on network nodes[48] or edges[41,58] or both[59]; or constrained scores by network topology[56] or signal content[44]; or are prioritized by high-scoring 'seed' nodes[60], including using strategies for computational colour coding of 'seed' paths[51,55]. Searching for active modules has proven to be a computationally difficult problem[48]. Hence, so-called heuristic solutions (for example, based on greedy[52,61–63], simulated annealing[48] or genetic[64] algorithms (BOX 1)) optimize computing time by recovering high-scoring subnetworks without necessarily finding the maximally scoring subnetworks. Nevertheless, exact methods that guarantee the detection of maximally scoring subnetworks, albeit at higher computational expense, have been programmed to run in fast timescales[44,45,65,66].

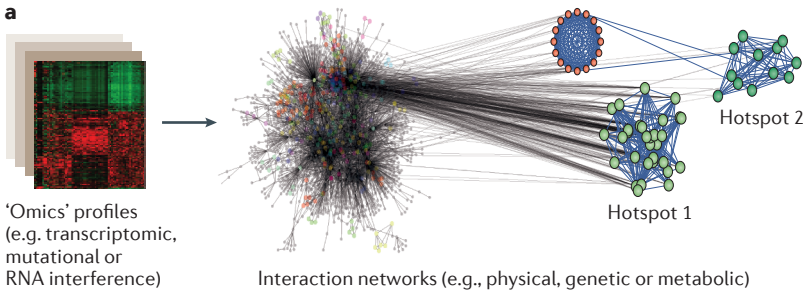
Diffusion-flow and network-propagation methods. The second group of methods for active module identification emulates the related concepts of diffusion flow and network propagation[36,37,45,67–72]. Analogous to fluid or heat flow through a system of pipes, network 'flow' is 'diffused' from nodes that are implicated in molecular profiles, such as differentially expressed genes or known disease genes. The flow reaches outwards along network edges, allowing subsequent identification of active modules as those subnetworks that accumulate maximum flow.

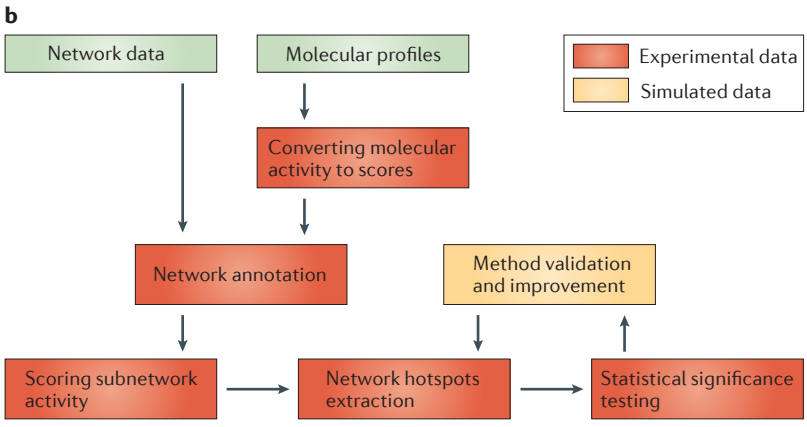
Figure 1 | **Identifying active modules. a** | Schematic representation of active modules inferred through the integration of biological networks and cellular state profiles. **b** | Common procedural workflow involved in active module identification. **c** | An active module of chromatin remodelling genes (highlighted in bold text) that were found to be mutated in clear-cell renal-cell carcinoma[72]. The module was identified through integrative analysis of multiple omics data sets using a combination of bioinformatics approaches (the TieDIE extension of HotNet, and PARADIGM). The module highlights several regulatory links connecting mutated genes to transcriptionally active targets. Each gene is depicted as a set of concentric rings representing various levels of biological information, and each 'spoke' in a ring pertains to a single patient sample. From the periphery inwards, the rings indicate PARADIGM-inferred levels of gene activity, mRNA expression levels, mutational abundance and correlation of gene expression or activity to mutation events in chromatin-related genes. Part **c** is reproduced, with permission, from REF. 72 © (2013) Macmillan Publishers Ltd. All rights reserved.

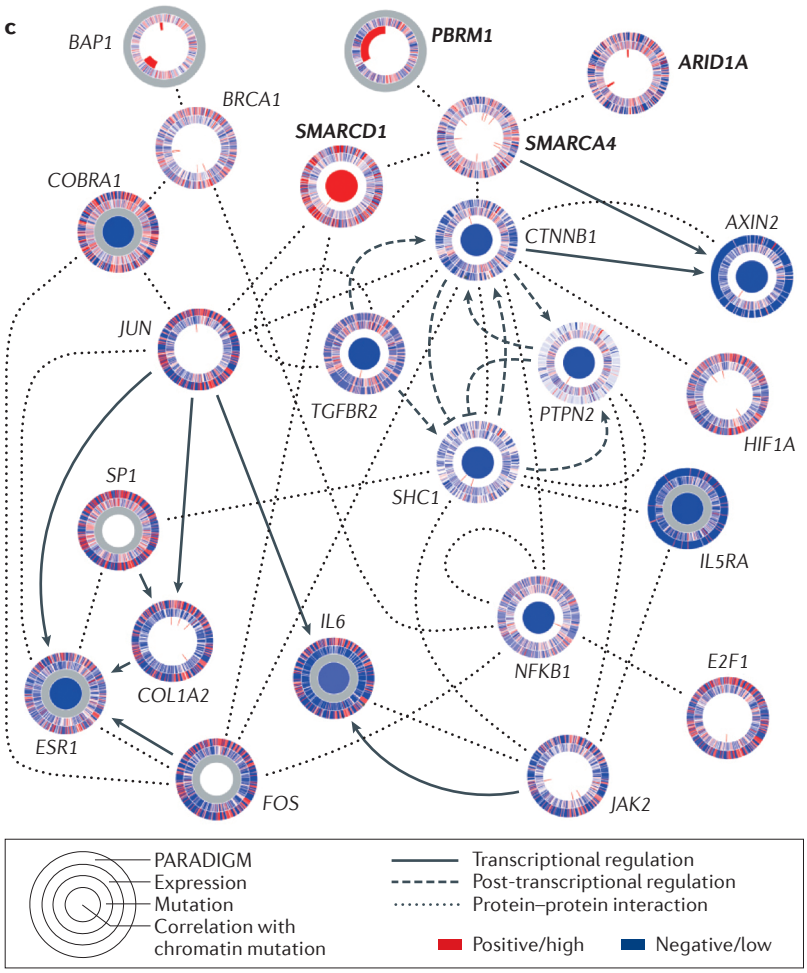
Degree
The number of interactions (edges) that a molecule (node) has in a network.

Betweenness centrality
A statistical intuition of how 'central' the status of a given molecule (node) or interaction (edge) is within a network. This is inferred by the fraction of shortest paths between all pairs of nodes that pass through a particular node or edge.

Network topology
The overall arrangement of nodes and edges in a given network.







Recently, network propagation methods such as those implemented in the bioinformatics tools HotNet and TieDIE (TABLE 1) have been used for network mapping of cancer mutations. These methods have proven particularly valuable for discovering mutational hotspots in human cancers[67,70–74]. For example, in one implementation of the application tool HotNet[67], significantly mutated pathways in glioblastomas and adenocarcinomas were identified through network propagation of associated cancer mutation profiles. Here, diffusion flow was run on a human protein–protein network that was seeded from known cancer genes to map their global neighbourhood of interaction. This operation translates to computing the 'influence' of cancer genes on all remaining genes in the network (BOX 1). The resulting 'influence network' (representing the full set of network connectivities surrounding cancer seed genes) was subsequently partitioned into weighted subnetworks. Thresholds were applied to these subnetworks according to either the number of patients in which they were mutated, or by the average number of somatic mutations that were associated per interacting gene pair in a given subnetwork, as informed by tumour sequence profiles. The highest weighted subnetworks marked significantly mutated cancer pathways. Such strategies have become increasingly popular and data-rich owing to easy availability of genome sequence and other 'omics profiles in public repositories such as The Cancer Genome Atlas (TCGA)[38,71,72].

Additionally, numerous propagation-based tools such as RegMod[45], ResponseNet[75] and NetWalker[76] (TABLE 1) permit functional network analysis informed by transcriptomic data. For example, ResponseNet traces information flow from upstream response regulators through signalling and regulatory pathways embedded in integrated protein networks to provide pathway-based explanations for downstream transcriptional changes that are captured in gene expression profiles.

Network propagation methods are particularly suitable for annotation, ranking or clustering of genes (such as disease genes) based on affiliations formed by network connectivity. In these situations, deciphering the precise architecture of a network is usually not a primary goal. Rather, the main motivation behind network propagation is to take advantage of the general functional proximity of genes to one another. Hence, the phrase 'network smoothing' is often used to describe such strategies.

Clustering-based methods. The third group of methods uses simultaneous clustering of network interactions and the conditions under which these interactions are active, in a concept termed 'biclustering'[46]. Clustering based on network topology alone has proven instrumental in defining basic principles of modular network organization[7,77,78]. Biclustering algorithms further expand these capabilities by evaluating both network connectivity and the correlation of omics-based performance across multiple samples or conditions[36,46,79,80]. A quantitative assessment of biclustering methods was recently

Metabolic flux
The flow of chemicals through any metabolic reaction (for example, an enzymatic reaction).

presented[46]. Many biclustering methods have been adapted as application tools (TABLE 1), such as SANDY[81], SAMBA[82] and cMonkey[69] (BOX 1). These tools permit multiplexed data analysis by interpreting global network topology and statistics in contexts of transcriptional regulatory information, differential expression profiles across multiple conditions and/or other types of biomedical information (such as phenotypic, sequence-based, literature and/or clinical information).

Modules derived through such a broad range of data types, covering multiple levels of biological regulation, are providing increasingly comprehensive interpretations of biological systems. For example, methods have also been developed for identifying active modules within metabolic networks, in which omics or regulatory data are used to constrain the allowable metabolic fluxes through the reactions in the network. High-flux reactions (edges) are clustered together and reported as

Table 1 | **Some recent bioinformatics tools for module extraction through network integration**

Tool	URL	Refs
Active-module detection through network projection of omics data		
jActiveModules	http://apps.cytoscape.org/apps/jactivemodules	48
MATISSE	http://acgt.cs.tau.ac.il/matisse	165
PinnacleZ	http://apps.cytoscape.org/apps/pinnaclez	62
GXNA	http://stat.stanford.edu/~serban/gxna	52
BioNet	http://bionet.bioapps.biozentrum.uni-wuerzburg.de	166
COSINE	http://cran.r-project.org/web/packages/COSINE/index.html	104
SANDY	http://sandy.topnet.gersteinlab.org	81
HotNet	http://ccmbweb.ccv.brown.edu/hotnet	67
PARADIGM	http://sbenz.github.com/Paradigm	70
MEMo	http://cbio.mskcc.org/memo	73
Multi-Dendrix	http://compbio.cs.brown.edu/software	37
RegMOD	http://www.biomedcentral.com/1471-2105/11/26/additional	45
NetWalk and FunWalk	http://netwalkersuite.org	76
ResponseNet	http://bioinfo.bgu.ac.il/respnet	75
ClustEx	http://www.mybiosoftware.com/pathway-analysis/5495	42
SAMBA	http://acgt.cs.tau.ac.il/samba	82
cMonkey	http://bonneaulab.bio.nyu.edu/biclustering.html	69
COBRAv2.0	http://opencobra.sourceforge.net/openCOBRA/Welcome.html	85
TieDIE	https://sysbiowiki.soe.ucsc.edu/tiedie	167
Network comparisons across species to identify conserved modules		
PathBLAST	http://www.pathblast.org	114
NetworkBLAST	http://www.cs.tau.ac.il/~bnet/networkblast.htm	168
NetworkBLAST-M	http://www.cs.tau.ac.il/~bnet/License-nbm.htm	116
IsoRankN	http://groups.csail.mit.edu/cb/mna	169
Graemlin	http://graemlin.stanford.edu	119
NeXus	http://csbio.cs.umn.edu/neXus/help.html	157
Multi-species cMonkey	http://bonneaulab.bio.nyu.edu/biclustering.html	158
Differential analysis of interaction networks to identify dynamic modules		
DDN	http://www.cbil.ece.vt.edu/software.htm	170
DNA	http://www.somnathdatta.org/Supp/DNA	171
Integration of diverse types of interaction networks to identify composite modules		
PanGIA	http://prosecco.ucsd.edu/PanGIA	147

BLAST, basic local alignment search tool; ClustEx, gene clustering and extending; COBRA, constraints-based reconstruction and analysis; COSINE, condition-specific subnetwork; DDN, differential dependence networks; DNA, differential network analysis (definition applies to this table only); Graemlin, general and robust alignment of multiple large interaction networks; GXNA, gene expression network analysis; MATISSE, module analysis via topology of interactions and similarity sets; MEMo, mutually exclusive modules in cancer; Multi-Dendrix, multiple pathway *de novo* driver exclusivity; NeXus, network cross(X)-species search; PanGIA, physical and genetic interaction alignment; PARADIGM, pathway recognition algorithm using data integration on genomic models; RegMOD, regression model with a diffusion kernel; SAMBA, statistical algorithmic method for bicluster analysis; SANDY, statistical analysis of network dynamics; TieDIE, tied diffusion through interacting events.

Box 1 | **Common bioinformatics themes applied in integrative module-finding approaches**

Simulated annealing
Simulated annealing is an optimization procedure that mimics the process undergone by misplaced atoms in a metal when it is heated and then slowly cooled. It was the first heuristic approach to be applied to the active module search problem[48]. To begin, a connected subgraph is chosen at random and scored as the average value of its nodes, taken from a molecular profiling experiment. Over many iterations, nodes are added or removed from this subgraph, and these changes are retained if they result in a connected subgraph with a better score. The changes may also be retained if they lower the score, with a probability that scales with the 'annealing temperature'. With each iteration, the temperature is lowered such that the accepted changes are increasingly likely to be beneficial. The final high-scoring subgraph is returned as the most 'active module'.

Greedy algorithms
Greedy algorithms are heuristic optimization algorithms that make the locally optimal choice at each stage. For example, in one greedy-based scheme[52], subnetworks were iteratively expanded from high-degree nodes either until the aggregate subnetwork score surpassed a predefined threshold or until the subnetwork size was saturated. Alternately, only nodes within a fixed radius of the seed node were aggregated[62]. In a greedy variant of simulated annealing, only a limited number of negative-scoring nodes (inactive nodes) was added in each iterative expansion step[61].

Genetic algorithms
Genetic algorithms mimic natural selection among members of a population and involve the iterative computation of various combinations of solutions; those with the best fitness scores are selected. In one hotspot-detection method based on genetic algorithms[64], node fitness was estimated using both molecular activity and network topology.

Exact methods
Exact methods are guaranteed to identify a maximally scoring subnetwork. They often have long run times although some have been made quite efficient[44,45,65,66]. One such method[44] allowed a rapid recovery of modules by transforming the subnetwork search task into a well-known prize-collecting Steiner trees (PCST) problem and solving it using integer linear programming (ILP).

Network propagation
Network propagation methods (also known as network smoothing methods) propagate network flow from selected nodes to identify subnetworks that accumulate the maximum 'flow' (that is, influence from neighbouring nodes). In one such method[67], an 'influence graph' was generated by releasing flow from cancer genes (that is, source (s)) along interaction edges. The influence graph was decomposed into component subnetworks of high network connectivity and activity (mutational frequency).

Biclustering methods
These methods allow the simultaneous clustering of interaction data and omics profiles to identify co-regulated or correlated modules. In a biclustering method, cMonkey[69], *P* values of correlated expression, sequence similarity and network topology were measured and an aggregate *P* value was defined as the joint membership probability. Using simulated annealing, nodes with high joint membership values (that is, ≈ 1) were iteratively aggregated; those with low values (that is, ≈ 0) were dropped; whereas those with intermediate values were added with decreasing probability per iteration (heat gradient) to identify hotspots.

active modules. We refer the reader to recent reviews[83,84] on integrative methods for modelling of metabolic networks through omics-based constraints. A version of the application tool COBRA (constraint-based reconstruction and analysis) (TABLE 1) permits omics-constrained analyses of genome-scale metabolic networks to predict feasible metabolic phenotypes and relevant modules under a given set of conditions[85].

Applications of active modules

Active modules have been identified using a wide range of interaction types (for example, protein–protein, regulatory and metabolic interactions) and 'omics' profiles (for example, gene expression profiles, mutation status data, RNA interference phenotypes and other cellular-state data), any combination of which may be applied for a single module-finding application. See Supplementary information S1 (table) for tools and databases related to various types of interactions and omics data.

A number of studies have interpreted omics profiles in the context of protein–protein interaction networks[34,39,48,50,62,67,70,72–74,80,86]. For example, recent work[72] established a comprehensive network view of molecular pathways altered in clear-cell renal-cell carcinoma (ccRCC) by analysing a diverse cohort of TCGA-derived omics data that included gene expression, genome mutation and methylation profiles in conjunction with human protein–protein interactions. The HotNet and PARADIGM methods were used to identify phosphoinositide 3-kinase (PI3K) pathways and SWI/SNF chromatin remodelling complexes as cancer-relevant active modules (FIG. 1c). Moreover, aberrant remodelling of cellular metabolism was repeatedly found to affect tumour stage and severity. Similarly, using the ResponseNet program, yeast networks of protein–protein, metabolic and protein–DNA interactions were analysed simultaneously with mRNA-profiling data to discover pathways that respond to α-synuclein toxicity[87].

Another study applied a method based on SigArSearch[62] (TABLE 1) to detect pathways of protein–protein interactions that show dysregulated expression in human breast cancer[62]. Compared with individual cancer-gene markers, these expression-based modules

showed greater accuracy in distinguishing metastatic from non-metastatic breast cancers, thus demonstrating the superior power of module-based biomarkers for disease prognosis. Alternatively, co-clustering of RNA interference data with protein–protein networks identified hepatitis C virus (HCV)-responsive modules in humans and established a role for the human ESCRT-III complex as an infection-permissive host factor[80]. Other discoveries of omics-derived modules using protein interaction knowledge have spanned a variety of model organisms, including metabolism in yeast[48], drug response in *Mycobacterium tuberculosis*[50], ageing in *Drosophila melanogaster*[88], ageing[56] and embryogenesis[34] in *Caenorhabditis elegans,* and cellular responses to inflammation[86], HIV infection[61] or tumour necrosis factor (TNF)-mediated stress[89] in humans.

Another prominent group of applications relates to the integration of omics profiles with protein–DNA interaction networks for the identification of active regulatory pathways[4,81,90]. For example, co-clustering of protein–DNA interactions and multi-condition gene expression profiles in yeast demonstrated widespread dynamic remodelling of transcription networks in response to diverse environmental stimuli[81]. It further showed that whereas a few transcriptional complexes act as constant 'hubs' of transcription, most appear transiently under particular conditions. In another study, differentially expressed arsenic-responsive pathways were extracted through the overlay of transcriptional profiles on yeast protein–DNA networks using the jActiveModules platform[90]. It was found that transcriptional data revealed important transcriptional complexes in gene-regulatory networks but not in metabolic networks, whereas phenotypic profiles (of arsenic sensitivity) mapped more cohesively onto metabolic networks.

The identification of active modules has also been applied to metabolic networks[50,90–92]. Constraint-based methods for analysing metabolic networks, including the widely exploited flux balance analysis (FBA) method, predict steady-state distributions of metabolic fluxes based on various physico-chemical constraints, such as rates of cellular growth and bioenergetics[93]. A recent variation on these methods adopts an integrative framework, whereby metabolic flux predictions are guided by omics or regulatory information (as reviewed in REFS 83,84). For example, a genome-scale reconstruction of a human metabolic network (curated from published data) was constrained using quantitative measures of gene and protein expression to predict tissue-specific metabolic uptake and release[91]. The study revealed a central role for post-transcriptional regulation in directing tissue-specific metabolic behaviours and associated metabolic diseases.

The discovery of active modules has paved the way for exciting diagnostic and therapeutic interventions. For example, active modules showing characteristic patterns of gene expression that correlate with specific disease phenotypes can yield valuable biomarkers for disease classification[62,94,95]. Module-based biomarkers achieve greater predictive power and reproducibility compared with single-gene biomarkers, as demonstrated for the classification of numerous human disorders including Alzheimer's disease[96], diabetes[36,97–99] and several forms of cancer including breast cancer[45,55,62,98,100,101], ovarian cancer[73,102,103], glioblastoma[67,70,73,74] and others[39,72,94,104,105]. Because active modules can reveal pathway-centric insights that are reinforced by multiple lines of evidence, they naturally provide mechanistic explanations for complex traits and polygenic diseases such as cancer. Moreover, active modules can assist in the discovery of drug-target pathways[50,106] and in predicting patient outcomes, such as response to chemotherapy[55].

Identification of conserved modules

Biological networks undergo substantial rewiring through evolutionary time, concomitant with gains, losses or modifications of gene functions[107–110]. Therefore, network modules showing conservation over large evolutionary distances are likely to reflect well-preserved 'core' functions that have been maintained by natural selection. Discovery of such 'conserved modules' can address fundamental questions about biological regulation while predicting evolutionary principles that shape network architectures. Some publicly available tools for finding conserved modules are summarized in TABLE 1.

Conserved interactions. In one of the most fundamental approaches to identifying conservation at the network level, individual interactions have been observed to occur between orthologous gene pairs in two species, corresponding to conserved protein–protein interactions (known as interologues)[111] or conserved regulatory interactions (known as regulogues)[112]. In one classic extension of this idea, a network of co-expressed gene pairs in humans, flies, worms and yeast was derived; subsequently, a clustering algorithm was used to extract conserved modules underlying cell cycle regulation and other core cellular processes[3].

Beyond conservation of individual interactions, comparisons of modules across species may reveal high overall consistency in structure and function despite a lack of one-to-one correspondence at the level of individual molecules or interactions. Hence, a group of approaches has been devised to align complex network structures, and these developments parallel the advances in computational solutions for cross-species sequence comparisons[113]. These 'network alignment' approaches can be organized as follows:

Pairwise network alignments. Computational methods for network alignment have greatly advanced evolutionary comparisons of network modules. For example, local network alignment tools such as PathBLAST[114] and NetworkBLAST[115] permit parallel comparisons of simple pathways (also known as linear paths) or subnetworks (also known as modules), respectively. These methods use a common heuristic workflow whereby a merged representation of two networks, denoted the 'network alignment graph', is searched for conserved

Hubs
Molecules with the highest number of interactions (degree) in a network.

Orthologous
Refers to the evolutionary relationship between two genes in two species that have descended from a common ancestor. Such genes are denoted as orthologues.

paths or subnetworks on the basis of a probabilistic log-likelihood model of interaction densities.

Parallel alignment of multiple networks. Network alignment has been progressively scaled for analysis of multiple (more than two) networks. For example, fast computation of conserved modules across as many as ten species was achieved in one study[116] by redefining the alignment graph in NetworkBLAST and treating multiple networks as separate layers that are linked through common orthology. Orthology, as in the above methods for identifying conserved modules, is commonly defined based on sequence homology. However, each gene or protein may potentially harbour multiple orthologues and paralogues owing to gene duplication events in any of the multiple species being compared. The resulting many-to-many correspondences between putative orthologues can introduce high computational complexity in network alignment methods, which can scale exponentially with the addition of each new species and corresponding network. To address this scalability issue when aligning graphs from multiple species, global alignment methods[117–119] identify functional orthologues on the basis of similar neighbourhood topologies across species (that is, the overall arrangement of interactions surrounding a gene or protein or molecule).

Network alignment incorporating evolutionary dynamics. An important question in network evolution pertains to how the evolutionary dynamics of genome alterations shape network architecture over time[120–122]. Network alignment methods for scoring module conservation such as MaWish[123] and others are increasingly incorporating evolutionary rates of gene deletion, insertion and/or duplication for an accurate representation of the network evolution model. One study[124] additionally accounted for the phylogenetic history of genes, through reconstruction of a conserved ancestral protein–protein interaction network (CAPPI) from multiple species and its subsequent projection onto the individual networks to identify conserved subnetworks across flies, worms and humans.

Applications of conserved modules. Conservation-based studies have provided fascinating insights into network evolution. For example, the identification of conserved metabolic genes and reactions across archea, bacteria and eukaryotes, followed by species clustering and simulations in the presence or absence of oxygen, indicated that the emergence of all three domains of life predated the widespread availability of atmospheric oxygen, and that adaptation to oxygen was coupled with increased network complexity and, concurrently, increased biological complexity[125].

Additionally, comparative analyses of conserved modules can supplement sequence-matching techniques for the prediction of function[113,126–129], on the basis of the premise that interaction partners of orthologous genes or proteins are also likely to be functionally conserved. This was illustrated in the proof-of-principle application of NetworkBLAST, in which thousands of previously uncharacterized protein functions were predicted on the basis of their conserved interaction neighbourhoods, which were inferred from the pairwise alignment of protein–protein interaction networks across yeast, worms and flies[115].

Evolutionary conservation can also support predictions of mechanisms of drug action: if a given drug targets elements of a module that is conserved across two evolutionarily distant model organisms, there is an increased probability that the same drug also targets the corresponding conserved module in humans[130]. Furthermore, the identification of evolutionarily diverged modules in pathogenic species can uncover pathogen-specific drug targets that are absent in humans[131].

Differential network modules

Molecular interactions can change dramatically in response to cellular cues, developmental stages, environmental stresses, pharmacological treatments and disease states[32,100,129,132,133]. Yet, the inherently dynamic wiring of molecular networks remains under-explored at the systems level, as interaction data are typically measured under single conditions (for example, standard laboratory growth media). Therefore, various so-called 'differential' network analyses (FIG. 2) have adopted an experimental approach whereby biological networks are measured and compared across conditions to identify interactions and modules that are differentially present, absent or modified.

Principles of differential network analyses. Analogous to differential expression analysis, differential network analysis involves pairwise subtraction of interactions that have been mapped in different experimental conditions[129]. The subtractive process filters out ubiquitous interactions (so-called 'housekeeping' interactions[129]) that are common to all static conditions of interest. By selectively extracting interactions that are relevant to the studied condition or phenotype, this reduces the typical complexity of static networks. Most notably, differential networks tap interaction spaces that are inaccessible to static networks. In particular, individual interactions that may be too weak (in terms of the magnitude of the interaction strength) to be captured in either static condition can be identified solely based on the significance of their differential measurements between the conditions[27,129]. Such differential interactions, once identified, may be further organized into modules using numerous hierarchical- or graph-clustering methods[47,134] or various Cytoscape[57]-based network analysis tools[135,136].

Applications. Physical networks assembled from quantitative protein–DNA and protein–protein binding data under different conditions were some of the first to be analysed in a differential mode. For example, using standard chromatin immunoprecipitation (ChIP)-based assays for protein–DNA interactions *in vivo* (Supplementary information S1 (table)), alterations in binding of transcription factors (TFs) to promoters

following amino acid starvation[10] or chemical induction of DNA damage[137] were mapped in yeast, thus providing insights into the dynamic regulation of stress response pathways. Similar comparisons of protein–protein interactions following epidermal growth factor (EGF) treatment in yeast have shed light on EGF-dependent signalling[138]. A recent study[139] that explored tissue-specific effects on network wiring demonstrated a profound role of tissue-regulated alternative splicing on dynamic remodelling of protein–protein interaction networks. Using a luminescence-based mammalian interactome mapping (LUMIER) approach for measuring physical binding between experimentally chosen 'bait' (seed) and 'prey' (target) proteins, the authors mapped protein–protein interactions between normally functioning prey proteins and several neurally regulated bait proteins. These bait proteins were genetically engineered to include or exclude specific exons with the purpose of exploring exon-dependent effects on network wiring in human cells. The study found that almost one-third of neurally regulated exons that were tested significantly modulated protein–protein interactions, and that overall, tissue-dependent exons participated in more protein–protein interactions than other proteins.

Differential analysis has also been carried out across functional networks (that is, as opposed to physical networks; see Supplementary information S1 (table)). For example, we applied an approach termed differential epistasis mapping (dE-MAP) to compare genetic networks that are induced by different types of DNA-damaging agents[27,140]. In another example, gene co-expression networks from transcriptomic profiles of normal cells and prostate cancer samples were compared to identify subnetworks that are induced in prostate cancer[141]. In this study, differential, but not static, networks successfully detected known prostate-cancer-specific interactions for RAD50 and telomeric repeat-binding factor 2 (TRF2).

Similarly, metabolic networks assembled from correlated activities of liver metabolites were differentially compared between normal and diabetic conditions to identify functional regulators of diabetic dyslipidemias in humans[142]. It is likely that continued advances in differential network mapping and analysis will shed light on tissue-specific, spatiotemporal and dosage-dependent rewiring of biological networks.

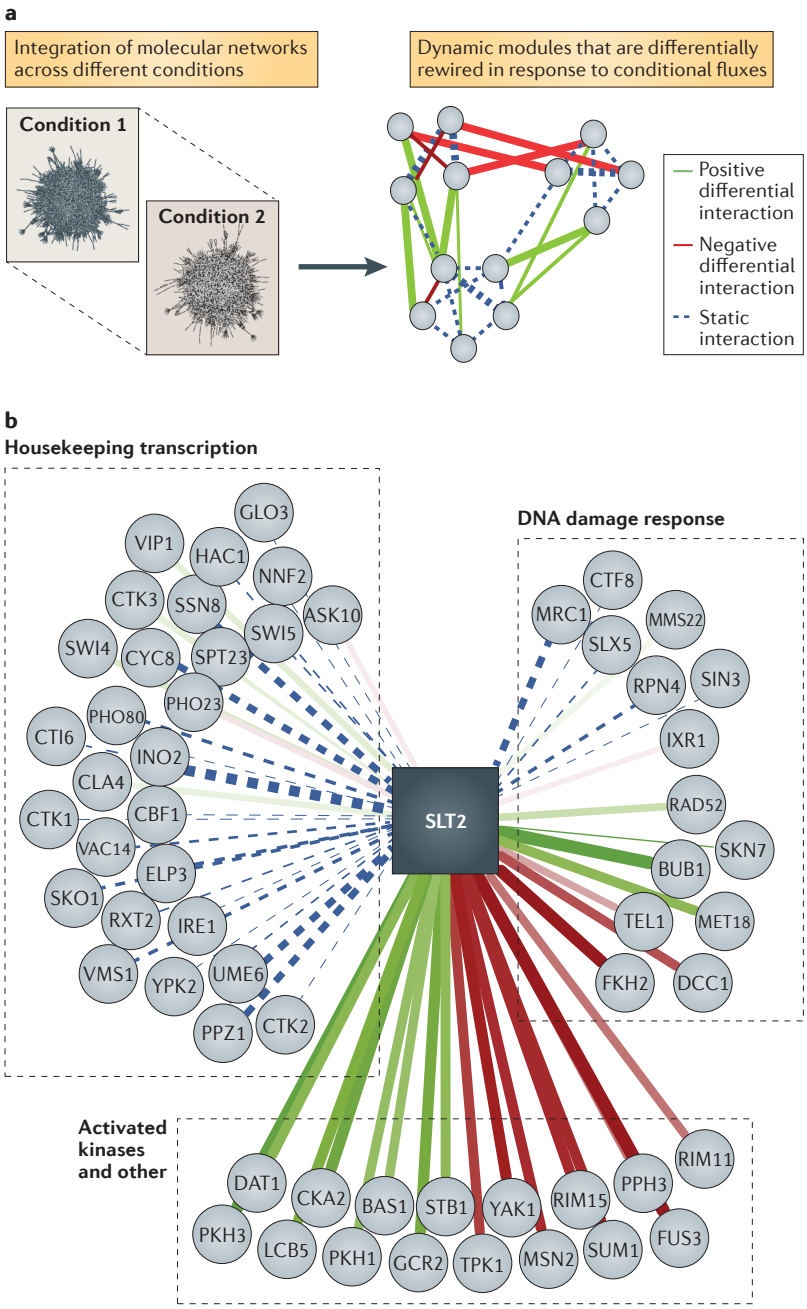


Figure 2 | **Differential analysis of molecular networks across conditions.** **a** | Schematic representation of a differential mapping approach to identify dynamically rewired modules. Molecular networks are assembled under multiple static conditions, and these static networks are subtracted across a pair of conditions to deduce differentially enriched interactions and modules. Edge widths indicate interaction strengths. **b** | Differential wiring of the genetic interactions of the serine/threonine mitogen-activated protein kinase SLT2 before and after methyl-methanesulphonate-induced DNA damage in yeast. Differential interactions effectively capture DNA damage response genes and several context-dependent kinases, whereas static interactions are enriched for housekeeping functions. Part **b** is modified from REF. 129.

Discovery of composite functional modules

Rationale for composite modules. Different types of biological interactions provide distinct, yet complementary, insights into cellular structure and function. For example, protein–protein, regulatory and metabolic networks each reflect a different aspect of the physical architecture of a cell (Supplementary information S1 (table)). Moreover, genetic interactions, which quantify epistatic effects of one gene on the phenotype expressed by another gene, reveal functional relationships between gene pairs. A key opportunity lies in reconciling these complementary network views of the cell into cohesive models. Powerful integrative approaches aimed at identifying composite functional modules that are composed of multiple types of biological interactions are providing considerable advances in this direction.

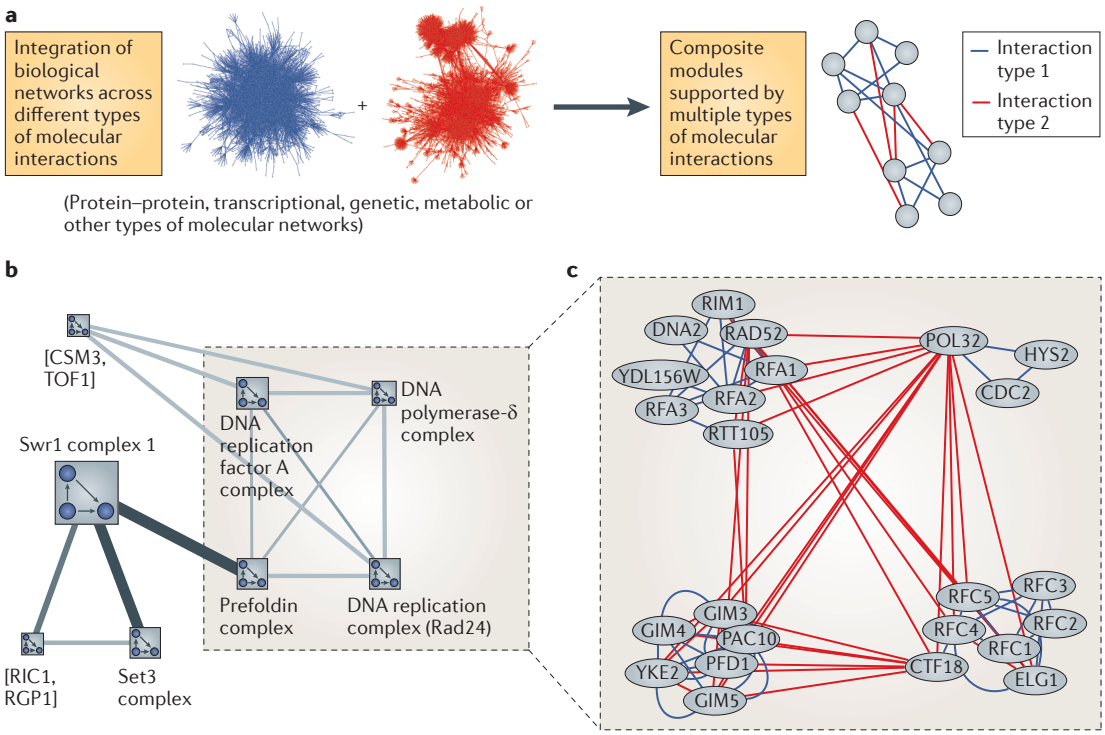


Figure 3 | **Integrating networks across interaction types.** **a** | A schematic view of composite functional modules identified through computational integration of diverse types of interaction networks. **b** | A hierarchical representation of modular and intermodule relationships inferred through joint analysis of physical (protein–protein) and genetic (epistatic) interactions (from supplementary data files in REF. 147) using the Cytoscape[57]-based application tool, PanGIA[147]. Here, module memberships are determined on the basis of physical and genetic interaction densities. Composite (physical–genetic) modules are represented as boxes, whereas edges between boxes represent the density of intermodule genetic interactions, that is, connecting genes across the two modules. **c** | A magnified internal view of four network modules. Blue edges represent protein–protein interactions and red edges represent genetic interactions within and between these modules, indicating intra-pathway and cross-pathway functional inter-dependencies, respectively.

Modes and applications. One class of approaches maps composite modules that are jointly supported by physical and genetic interactions[143] (FIG. 3). A common theme in these approaches[13,128,144–146], implemented in the PanGIA application[147] (TABLE 1), involves the identification of overlapping clusters of physical and genetic interactions; these composite modules implicate genes acting within a pathway. Clusters of genetic interactions that bridge two different composite modules reflect intermodule dependencies that link synergistic, compensatory or redundant pathways[144]. Integrative analysis of composite modules encompassing physical and genetic interactions can reveal physical mechanisms underlying phenotypes caused by mutations from genetic screens or, conversely, can predict genetic dependencies between protein complexes that have been mapped in physical binding assays. Module maps elucidating global physical–genetic inter-relations have been assembled in a number of studies exploring heat shock protein 90 (Hsp90) signalling[148], chromosomal biology[13,145], RNA processing[149], secretory pathways[150], DNA damage responses[27] or global biological processes[144,151].

Integrative strategies have similarly uncovered composite modules in signalling and regulatory networks, primarily through combined evaluation of protein–DNA interactions (specifically, TF–target interactions) and protein–protein interactions[11,59,152,153], or by additionally including genetic interactions[151]. In early work along these lines, composite 'motifs' comprised of regulatory and protein–protein interactions among 2, 3 or 4 proteins were mapped and classified into distinct feed-forward loops, interacting transcriptional hubs and other logical circuits[152]. Such simple motifs were thought to combine with recurrent patterns to organize higher-order network 'themes'; that is, complex functional modules that are associated with specific biological responses[151]. In a related study[153], yeast protein–protein and protein–DNA interaction networks were combined to identify 72 co-regulated protein complexes. Such co-regulated complexes depict dense protein clusters (in protein–protein networks) for which members are jointly regulated by a common set of transcription factors (in corresponding protein–DNA networks). At the network level, these TF–protein co-complexes were visualized along with their regulatory relationships to the other (non-transcriptional) modules that they regulate. An evolutionary comparison of these co-regulated complexes suggested that

protein complexes may evolve with slower dynamics than protein–DNA transcriptional relationships. Related studies exploring co-regulated complexes in yeast have revealed cross-pathway communication between hyperosmotic, heat shock and oxidative stress response systems[59], and have elucidated signalling networks that are active during pheromone responses[53].

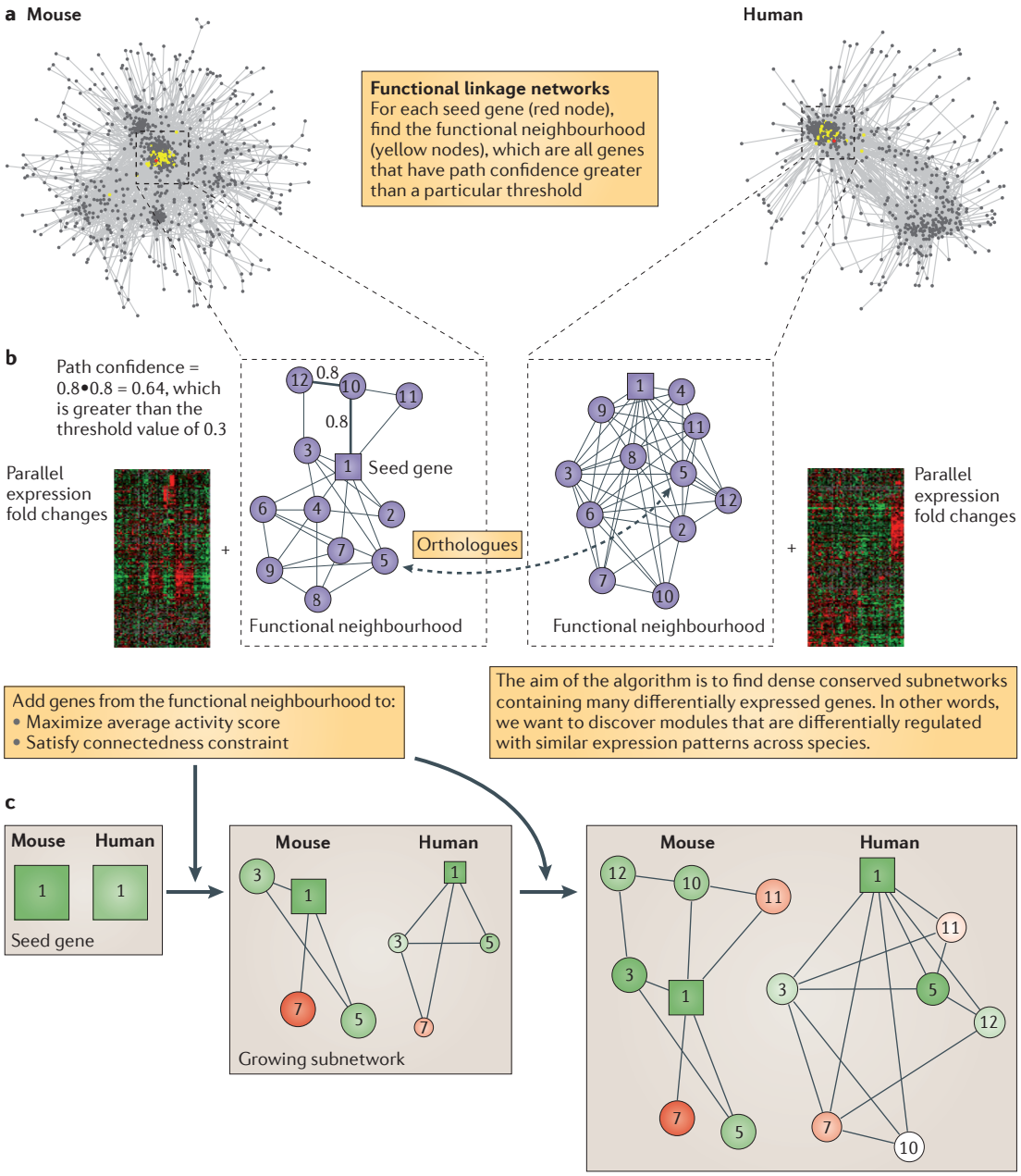


Figure 4 | **Identification of conserved functional modules by integration of data across multiple species.** **a** | Functional linkage networks are assembled from multiple lines of evidence (for example, protein–protein and genetic interactions, gene expression, protein localization, phenotypes and sequence data) and integrated with differential gene expression profiles. This example is derived from human and mouse tissues (stem cells and differentiated cells). Candidate seed genes (red) are defined as differentially expressed orthologues. The functional neighbourhood (yellow) of each seed gene is marked by genes for which path confidence (the product of linkage weights along the path) from the seed gene exceeds a specified threshold. **b** | A search for modules seeks densely connected subnetworks of genes that share similar patterns of expression in both species. **c** | In this search, subnetworks are grown simultaneously in both species starting from the seed genes ('1' in the square) and expanded through the iterative addition of genes that satisfy both of two criteria: first, the genes must be in the same functional neighbourhood, and second, the genes must maximize a differential expression activity score. Differentially expressed genes are coloured green (upregulated) or red (downregulated). The figure is modified from REF. 157.

Protein–DNA interactions have also been combined with metabolic networks to understand the effects of transcriptional regulation on biochemical output[83,84,90,154–156]. For example, the probabilistic regulation of metabolism (PROM) method was developed to facilitate automated and quantitative integration of regulatory interactions and other high-throughput data for constraint-based modelling of metabolic networks[156]. The method was applied for a genome-scale analysis of an integrative metabolic regulatory network model for *M. tuberculosis*, incorporating information from greater than 2,000 TF–gene promoter interactions regulating 3,300 metabolic reactions, 1,300 expression profiles and 1,905 deletion phenotypes from *E. coli* and *M. tuberculosis*. The method enabled a powerful prediction of microbial growth phenotypes under various environmental perturbations and aided the identification of novel gene functions. Furthermore, the study isolated several transcription factor hubs that regulate multiple target proteins in the pathogen interactome as a strategy for uncovering promising antimicrobial drug targets.

Combined application of integrative approaches

Given the above four integrative approaches, a very recent trend has been to chain together more than one of these approaches to create network analysis pipelines of increasing sophistication and complexity. For example, network-module-finding methods based on integration across molecular profiles and network types (for example, for finding active modules or composite modules) have been extended across species for extracting co-functional modules that are also conserved. A multi-species and scalable framework, neXus (network cross(X)-species search)[157], was developed for discovering conserved functional modules through parallel expression profiling in multiple species (FIG. 4). Specifically, a clustering-based approach was used to extract subnetworks from functional linkage networks (incorporating a wide range of interaction and omics information) that had been derived independently from mouse and human samples. Subnetworks were seeded from differentially expressed orthologues, and then expanded simultaneously for both species. Using programmatic constraints to apply thresholds to candidate subnetworks according to network connectivity and molecular activity, conserved active subnetworks were nominated. These subnetworks showed significant differential activity in stem cells relative to differentiated cells and shared similar patterns of gene expression across mouse and human samples. An extended version of the cMonkey framework that was designed for simultaneous (rather than sequential) data integration across multiple species[158] (TABLE 1) further expands the scope of such analyses. It allows parallel evaluation of protein–protein interactions, transcriptomic data, sequence profiles, metabolic and signalling pathway models and comparative genomics from multiple species to infer conserved co-regulated modules.

Another recent study[159] mapped global genetic networks in the fission yeast *Schizosaccharomyces pombe* and compared them with integrated maps of existing genetic and protein–protein networks (composite modules) in the divergent budding yeast *Saccharomyces cerevisiae*, with the aim of identifying conserved functional modules. The authors demonstrated a hierarchical model for the evolution of genetic interactions: interactions among genes that encoded proteins in the same protein complex showed the highest degree of conservation, those involved within the same biological process showed lower but still significant conservation, whereas those participating in different biological processes were poorly conserved. Conservation of cross-pathway interactions between distinct biological processes was observed on a larger scale. Together, these observations reveal functional and evolutionary design principles underlying the modular organization of cellular networks.

With continued progress in integrative bioinformatics pipelines and the expansion of data-handling capabilities, a very large combination of data types, conditions, species, time points and cell states could potentially be amenable to joint evaluation for in-depth network analysis.

Perspective

The past decade has witnessed an explosive growth in data on biological networks[9–14,16,160,161], albeit with inherent limitations[24] and largely from a static perspective[129]. The integrative approaches reviewed here substantially increase the scope, scale and depth of network analyses, by permitting joint interpretation of ensembles of biological information. Although these strategies have greatly refined high-throughput data analysis by tackling several of its prevalent challenges — such as variability in accuracy and coverage, and context-specificity — even greater power for mining biological knowledge remains to be achieved by implementing a combination of these approaches. Such combination strategies that encompass multiple algorithms, data types, conditions and species contexts are likely to maximize the performance, relevance and scope of module-assisted network analysis. For example, although it has not yet been attempted, it would be conceivable to analyse differential networks (discussed above in the 'Differential network modules' section) across multiple species (discussed above in the 'Identification of conserved modules' section) to detect conserved dynamic modules and process-specific pathways. Another challenging direction would be to study the evolution of composite modules, as it is becoming increasingly clear that different network types exhibit specific evolutionary dynamics; for example, regulatory interactions evolve more rapidly than genetic, protein and metabolic networks[162].

Module-based biomarkers derived through integrative network analyses also provide superior predictive performance in disease classification, especially when compared with single-gene disease markers that have been routinely annotated through genome-wide association studies (GWASs)[38,62,71,72,163,164]. Future work on integrative network analyses will provide greater insight into pathway structures and highlight network-level dynamics underlying biological responses.

1. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
2. Alon, U. Biological networks: the tinkerer as an engineer. *Science* **301**, 1866–1867 (2003).
3. Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–255 (2003).
4. Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genet.* **34**, 166–176 (2003).
5. Barabasi, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature Rev. Genet.* **12**, 56–68 (2011).
6. Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet.* **5**, 101–113 (2004).
7. Spirin, V. & Mirny, L. A. Protein complexes and functional modules in molecular networks. *Proc. Natl Acad. Sci. USA* **100**, 12123–12128 (2003).
8. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
9. Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
10. Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
11. Ravasi, T. *et al.* An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744–752 (2010).
12. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
13. Collins, S. R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).
14. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
15. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
16. Muers, M. Systems biology: plant networks. *Nature Rev. Genet.* **12**, 586 (2011).
17. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
18. Ideker, T. & Sharan, R. Protein networks in disease. *Genome Res.* **18**, 644–652 (2008).
19. Koyuturk, M. Algorithmic and analytical methods in network biology. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 277–292 (2010).
20. Fields, S. High-throughput two-hybrid analysis. The promise and the peril. *FEBS J.* **272**, 5391–5399 (2005).
21. Phizicky, E. M. & Fields, S. Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59**, 94–123 (1995).
22. Ben-Hur, A. & Noble, W. S. Kernel methods for predicting protein-protein interactions. *Bioinformatics* **21** (Suppl. 1), i38–i46 (2005).
23. Huang, H., Jedynak, B. M. & Bader, J. S. Where have all the interactions gone? Estimating the coverage of two-hybrid protein interaction maps. *PLoS Comput. Biol.* **3**, e214 (2007).
24. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nature Methods* **6**, 83–90 (2009).
A critical discussion of biases in high-throughput data analyses that contribute to false positives and negative interpretations.
25. Cusick, M. E. *et al.* Literature-curated protein interaction datasets. *Nature Methods* **6**, 39–46 (2009).
26. Edwards, A. M. *et al.* Too many roads not taken. *Nature* **470**, 163–165 (2011).
27. Bandyopadhyay, S. *et al.* Rewiring of genetic networks in response to DNA damage. *Science* **330**, 1385–1389 (2010).
An approach for differential analysis of genetic networks. It was applied to the mapping of DNA damage response pathways in yeast.
28. Califano, A. Rewiring makes the difference. *Mol. Syst. Biol.* **7**, 463 (2011).
29. Ideker, T & Bandyopadhyay, S. Integrative systems biology poster [online] http://www.nature.com/ng/extra/sysbio. *Nature Genet.* **42** (2010).
30. Jenssen, T. K., Laegreid, A., Komorowski, J. & Hovig, E. A literature network of human genes for high-throughput analysis of gene expression. *Nature Genet.* **28**, 21–28 (2001).
31. Jansen, R., Greenbaum, D. & Gerstein, M. Relating whole-genome expression data with protein-protein interactions. *Genome Res.* **12**, 37–46 (2002).
32. de Lichtenberg, U., Jensen, L. J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **307**, 724–727 (2005).
33. Segal, E., Wang, H. & Koller, D. Discovering molecular pathways from protein interaction and gene expression data. *Bioinformatics* **19** (Suppl. 1), i264–i271 (2003).
34. Gunsalus, K. C. *et al.* Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* **436**, 861–865 (2005).
35. Jensen, L. J., Jensen, T. S., de Lichtenberg, U., Brunak, S. & Bork, P. Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature* **443**, 594–597 (2006).
36. Chen, R. *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148**, 1293–1307 (2012).
37. Leiserson, M. D., Blokh, D., Sharan, R. & Raphael, B. J. Simultaneous identification of multiple driver pathways in cancer. *PLoS Comput. Biol.* **9**, e1003054 (2013).
38. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
39. Nibbe, R. K., Koyuturk, M. & Chance, M. R. An integrative -omics approach to identify functional sub-networks in human colorectal cancer. *PLoS Comput. Biol.* **6**, e1000639 (2010).
40. Begley, T. J., Rosenbach, A. S., Ideker, T. & Samson, L. D. Hot spots for modulating toxicity identified by genomic phenotyping and localization mapping. *Mol. Cell* **16**, 117–125 (2004).
41. Guo, Z. *et al.* Edge-based scoring and searching method for identifying condition-responsive protein-protein interaction sub-network. *Bioinformatics* **23**, 2121–2128 (2007).
42. Gu, J., Chen, Y., Li, S. & Li, Y. Identification of responsive gene modules by network-based gene clustering and extending: application to inflammation and angiogenesis. *BMC Syst. Biol.* **4**, 47 (2010).
43. Wu, Z., Zhao, X. & Chen, L. Identifying responsive functional modules from protein-protein interaction network. *Mol. Cells* **27**, 271–277 (2009).
44. Dittrich, M. T., Klau, G. W., Rosenwald, A., Dandekar, T. & Muller, T. Identifying functional modules in protein-protein interaction networks: an integrated exact approach. *Bioinformatics* **24**, i223–i231 (2008).
This article describes a programmatically efficient scheme for module detection that bypasses inherent computational complexities underlying the extraction of high-confidence (that is, maximally scoring) subnetworks from omics data.
45. Qiu, Y. Q., Zhang, S., Zhang, X. S. & Chen, L. Detecting disease associated modules and prioritizing active genes based on high throughput data. *BMC Bioinformatics* **11**, 26 (2010).
46. Prelic, A. *et al.* A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics* **22**, 1122–1129 (2006).
47. Sharan, R., Ulitsky, I. & Shamir, R. Network-based prediction of protein function. *Mol. Syst. Biol.* **3**, 88 (2007).
48. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A. F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18** (Suppl. 1), S233–S240 (2002).
49. Sohler, F., Hanisch, D. & Zimmer, R. New methods for joint analysis of biological networks and expression data. *Bioinformatics* **20**, 1517–1521 (2004).
50. Cabusora, L., Sutton, E., Fulmer, A. & Forst, C. V. Differential network expression during drug and stress response. *Bioinformatics* **21**, 2898–2905 (2005).
51. Scott, J., Ideker, T., Karp, R. M. & Sharan, R. Efficient algorithms for detecting signaling pathways in protein interaction networks. *J. Comput. Biol.* **13**, 133–144 (2006).
52. Nacu, S., Critchley-Thorne, R., Lee, P. & Holmes, S. Gene expression network analysis and applications to immunology. *Bioinformatics* **23**, 850–858 (2007).
53. Huang, S. S. & Fraenkel, E. Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Sci. Signal.* **2**, ra40 (2009).
54. Chowdhury, S. A. & Koyutürk, M. Identification of coordinately dysregulated subnetworks in complex phenotypes. *Pac. Symp. Biocomput.* **2010**, 133–144 (2010).
55. Dao, P. *et al.* Optimally discriminative subnetwork markers predict response to chemotherapy. *Bioinformatics* **27**, i205–i213 (2011).
56. Fortney, K., Kotlyar, M. & Jurisica, I. Inferring the functions of longevity genes with modular subnetwork biomarkers of *Caenorhabditis elegans* aging. *Genome Biol.* **11**, R13 (2010).
57. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
58. Ulitsky, I. & Shamir, R. Identifying functional modules using expression profiles and confidence-scored protein interactions. *Bioinformatics* **25**, 1158–1164 (2009).
59. Wang, Y. C. & Chen, B. S. Integrated cellular network of transcription regulations and protein-protein interactions. *BMC Syst. Biol.* **4**, 20 (2010).
60. Breitling, R., Amtmann, A. & Herzyk, P. Graph-based iterative Group Analysis enhances microarray interpretation. *BMC Bioinformatics* **5**, 100 (2004).
61. Rajagopalan, D. & Agarwal, P. Inferring pathways from gene lists using a literature-derived network of biological relationships. *Bioinformatics* **21**, 788–793 (2005).
62. Chuang, H. Y., Lee, E., Liu, Y. T., Lee, D. & Ideker, T. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* **3**, 140 (2007).
63. Hwang, T. & Park, T. Identification of differentially expressed subnetworks based on multivariate ANOVA. *BMC Bioinformatics* **10**, 128 (2009).
64. Klammer, M., Godl, K., Tebbe, A. & Schaab, C. Identifying differentially regulated subnetworks from phosphoproteomic data. *BMC Bioinformatics* **11**, 351 (2010).
65. Zhao, X. M., Wang, R. S., Chen, L. & Aihara, K. Uncovering signal transduction networks from high-throughput data by integer linear programming. *Nucleic Acids Res.* **36**, e48 (2008).
66. Backes, C. *et al.* An integer linear programming approach for finding deregulated subgraphs in regulatory networks. *Nucleic Acids Res.* **40**, e43 (2012).
67. Vandin, F., Upfal, E. & Raphael, B. J. Algorithms for detecting significantly mutated pathways in cancer. *J. Comput. Biol.* **18**, 507–522 (2011).
68. Komurov, K., White, M. A. & Ram, P. T. Use of data-biased random walks on graphs for the retrieval of context-specific networks from genomic data. *PLoS Comput. Biol.* **6**, e1000889 (2010).
69. Reiss, D. J., Baliga, N. S. & Bonneau, R. Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. *BMC Bioinformatics* **7**, 280 (2006).
70. Vaske, C. J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26**, i237–i245 (2010).
71. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
72. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
A genome-scale effort that mapped significantly mutated pathways in human cancer through network projection of mutational profiles, leading to the identification of novel disease mechanisms.
73. Ciriello, G., Cerami, E., Sander, C. & Schultz, N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* **22**, 398–406 (2012).
74. Miller, C. A., Settle, S. H., Sulman, E. P., Aldape, K. D. & Milosavljevic, A. Discovering functional modules by identifying recurrent and mutually exclusive mutational patterns in tumors. *BMC Med. Genom.* **4**, 34 (2011).
75. Lan, A. *et al.* ResponseNet: revealing signaling and regulatory networks linking genetic and transcriptomic screening data. *Nucleic Acids Res.* **39**, W424–W429 (2011).
76. Komurov, K., Dursun, S., Erdin, S. & Ram, P. T. NetWalker: a contextual network analysis tool for functional genomics. *BMC Genomics* **13**, 282 (2012).
77. Rives, A. W. & Galitski, T. Modular organization of cellular networks. *Proc. Natl Acad. Sci. USA* **100**, 1128–1133 (2003).
78. Ravasz, E. & Barabasi, A. L. Hierarchical organization in complex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **67**, 026112 (2003).
79. Hanisch, D., Zien, A., Zimmer, R. & Lengauer, T. Co-clustering of biological networks and gene expression data. *Bioinformatics* **18** (Suppl. 1), S145–S154 (2002).

80. Gonzalez, O. & Zimmer, R. Contextual analysis of RNAi-based functional screens using interaction networks. *Bioinformatics* **27**, 2707–2713 (2011).
81. Luscombe, N. M. *et al.* Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308–312 (2004).
This article presents an omics-based mapping of stress-response pathways in yeast protein networks, revealing key regulatory insights into network dynamics.
82. Tanay, A., Sharan, R., Kupiec, M. & Shamir, R. Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. *Proc. Natl Acad. Sci. USA* **101**, 2981–2986 (2004).
This study develops and integrates a widely cited approach for module discovery that allows the simultaneous interpretation of a diverse range of biological information.
83. Blazier, A. S. & Papin, J. A. Integration of expression data in genome-scale metabolic network reconstructions. *Front. Physiol.* **3**, 299 (2012).
84. Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nature Rev. Microbiol.* **10**, 291–305 (2012).
85. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protoc.* **6**, 1290–1307 (2011).
86. Calvano, S. E. *et al.* A network-based analysis of systemic inflammation in humans. *Nature* **437**, 1032–1037 (2005).
87. Yeger-Lotem, E. *et al.* Bridging high-throughput genetic and transcriptional data reveals cellular responses to α–synuclein toxicity. *Nature Genet.* **41**, 316–323 (2009).
88. Xue, H. *et al.* A modular network model of aging. *Mol. Syst. Biol.* **3**, 147 (2007).
89. Bandyopadhyay, S., Kelley, R. & Ideker, T. Discovering regulated networks during HIV-1 latency and reactivation. *Pac. Symp. Biocomput.* **2006**, 354–366 (2006).
90. Haugen, A. C. *et al.* Integrating phenotypic and expression profiles to map arsenic-response networks. *Genome Biol.* **5**, R95 (2004).
91. Shlomi, T., Cabili, M. N., Herrgard, M. J., Palsson, B. O. & Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nature Biotech.* **26**, 1003–1010 (2008).
92. Colijn, C. *et al.* Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput. Biol.* **5**, e1000489 (2009).
93. Price, N. D., Reed, J. L. & Palsson, B. O. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Rev. Microbiol.* **2**, 886–897 (2004).
94. Chowdhury, S. A., Nibbe, R. K., Chance, M. R. & Koyuturk, M. Subnetwork state functions define dysregulated subnetworks in cancer. *J. Comput. Biol.* **18**, 263–281 (2011).
95. Anastassiou, D. Computational analysis of the synergy among multiple interacting genes. *Mol. Syst. Biol.* **3**, 83 (2007).
96. Ma, X., Lee, H., Wang, L. & Sun, F. CGI: a new approach for prioritizing genes by combining gene expression and protein-protein interaction data. *Bioinformatics* **23**, 215–221 (2007).
97. Li, W. *et al.* Dynamical systems for discovering protein complexes and functional modules from biological networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **4**, 233–250 (2007).
98. Yang, P., Li, X., Wu, M., Kwoh, C. K. & Ng, S. K. Inferring gene-phenotype associations via global protein complex network propagation. *PLoS ONE* **6**, e21502 (2011).
99. Tu, Z. *et al.* Integrating siRNA and protein-protein interaction data to identify an expanded insulin signaling network. *Genome Res.* **19**, 1057–1067 (2009).
100. Taylor, I. W. *et al.* Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotech.* **27**, 199–204 (2009).
An omics-based strategy for identifying breast cancer pathways, which demonstrated the power of integrative network analysis for disease prognosis.
101. Zhang, X. *et al.* The expanded human disease network combining protein-protein interaction information. *Eur. J. Hum. Genet.* **19**, 783–788 (2011).
102. Bapat, S. A., Krishnan, A., Ghanate, A. D., Kusumbe, A. P. & Kalra, R. S. Gene expression: protein interaction systems network modeling identifies transformation-associated molecules and pathways in ovarian cancer. *Cancer Res.* **70**, 4809–4819 (2010).
103. Zhang, K. X. & Ouellette, B. F. CAERUS: predicting CAncER oUtcomeS using relationship between protein structural information, protein networks, gene expression data, and mutation data. *PLoS Comput. Biol.* **7**, e1001114 (2011).
104. Ma, H., Schadt, E. E., Kaplan, L. M. & Zhao, H. COSINE: COndition-SpecIfic sub-NEtwork identification using a global optimization method. *Bioinformatics* **27**, 1290–1298 (2011).
105. Ahn, J., Yoon, Y., Park, C., Shin, E. & Park, S. Integrative gene network construction for predicting a set of complementary prostate cancer genes. *Bioinformatics* **27**, 1846–1853 (2011).
106. Wu, Z., Zhao, X. M. & Chen, L. A systems biology approach to identify effective cocktail drugs. *BMC Syst. Biol.* **4**, (Suppl. 2), S7 (2010).
107. Vespignani, A. Evolution thinks modular. *Nature Genet.* **35**, 118–119 (2003).
108. Mazurie, A., Bonchev, D., Schwikowski, B. & Buck, G. A. Evolution of metabolic network organization. *BMC Syst. Biol.* **4**, 59 (2010).
109. Odom, D. T. *et al.* Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature Genet.* **39**, 730–732 (2007).
110. Wuchty, S., Oltvai, Z. N. & Barabasi, A. L. Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nature Genet.* **35**, 176–179 (2003).
111. Matthews, L. R. *et al.* Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". *Genome Res.* **11**, 2120–2126 (2001).
112. Yu, H. *et al.* Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res.* **14**, 1107–1118 (2004).
113. Sharan, R. & Ideker, T. Modeling cellular machinery through biological network comparison. *Nature Biotech.* **24**, 427–433 (2006).
114. Kelley, B. P. *et al.* PathBLAST: a tool for alignment of protein interaction networks. *Nucleic Acids Res.* **32**, W83–W88 (2004).
115. Sharan, R. *et al.* Conserved patterns of protein interaction in multiple species. *Proc. Natl Acad. Sci. USA* **102**, 1974–1979 (2005).
This study highlights a method for the pairwise alignment of subnetworks to facilitate efficient comparisons between diverse interactomes.
116. Kalaev, M., Bafna, V. & Sharan, R. Fast and accurate alignment of multiple protein networks. *J. Comput. Biol.* **16**, 989–999 (2009).
117. Bandyopadhyay, S., Sharan, R. & Ideker, T. Systematic identification of functional orthologs based on protein network comparison. *Genome Res.* **16**, 428–435 (2006).
118. Singh, R., Xu, J. & Berger, B. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. Natl Acad. Sci. USA* **105**, 12763–12768 (2008).
119. Flannick, J., Novak, A., Srinivasan, B. S., McAdams, H. H. & Batzoglou, S. Graemlin: general and robust alignment of multiple large interaction networks. *Genome Res.* **16**, 1169–1181 (2006).
120. Berg, J. & Lassig, M. Cross-species analysis of biological networks by Bayesian alignment. *Proc. Natl Acad. Sci. USA* **103**, 10967–10972 (2006).
121. Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
122. Arabidopsis Interactome Mapping Consortium. Evidence for network evolution in an *Arabidopsis* interactome map. *Science* **333**, 601–607 (2011).
123. Koyuturk, M. *et al.* Pairwise alignment of protein interaction networks. *J. Comput. Biol.* **13**, 182–199 (2006).
124. Dutkowski, J. & Tiuryn, J. Identification of functional modules from conserved ancestral protein-protein interactions. *Bioinformatics* **23**, i149–i158 (2007).
125. Raymond, J. & Segre, D. The effect of oxygen on biochemical networks and the evolution of complex life. *Science* **311**, 1764–1767 (2006).
126. Vazquez, A., Flammini, A., Maritan, A. & Vespignani, A. Global protein function prediction from protein-protein interaction networks. *Nature Biotech.* **21**, 697–700 (2003).
127. Sharan, R., Ideker, T., Kelley, B., Shamir, R. & Karp, R. M. Identification of protein complexes by comparative analysis of yeast and bacterial protein interaction data. *J. Comput. Biol.* **12**, 835–846 (2005).
128. Ulitsky, I. & Shamir, R. Pathway redundancy and protein essentiality revealed in the *Saccharomyces cerevisiae* interaction networks. *Mol. Syst. Biol.* **3**, 104 (2007).
129. Ideker, T. & Krogan, N. J. Differential network biology. *Mol. Syst. Biol.* **8**, 565 (2012).
130. Kapitzky, L. *et al.* Cross-species chemogenomic profiling reveals evolutionarily conserved drug mode of action. *Mol. Syst. Biol.* **6**, 451 (2010).
131. Suthram, S., Sittler, T. & Ideker, T. The *Plasmodium* protein network diverges from those of other eukaryotes. *Nature* **438**, 108–112 (2005).
132. Hillenmeyer, M. E. *et al.* The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–365 (2008).
133. Han, J. D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
134. Andreopoulos, B., An, A., Wang, X. & Schroeder, M. A roadmap of clustering algorithms: finding a match for a biomedical application. *Brief Bioinform.* **10**, 297–314 (2009).
135. Saito, R. *et al.* A travel guide to Cytoscape plugins. *Nature Methods* **9**, 1069–1076 (2012).
136. Yosef, N. *et al.* ANAT: a tool for constructing and analyzing functional protein networks. *Sci. Signal.* **4**, pl1 (2011).
137. Workman, C. T. *et al.* A systems approach to mapping DNA damage response pathways. *Science* **312**, 1054–1059 (2006).
138. Bisson, N. *et al.* Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nature Biotech.* **29**, 653–658 (2011).
139. Ellis, J. D. *et al.* Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* **46**, 884–892 (2012).
140. Guenole, A. *et al.* Dissection of DNA damage responses using multiconditional genetic interaction maps. *Mol. Cell* **49**, 346–358 (2013).
141. Altay, G., Asim, M., Markowetz, F. & Neal, D. E. Differential C3NET reveals disease networks of direct physical interactions. *BMC Bioinformatics* **12**, 296 (2011).
142. Valcarcel, B. *et al.* A differential network approach to exploring differences between biological states: an application to prediabetes. *PLoS ONE* **6**, e24702 (2011).
143. Beyer, A., Bandyopadhyay, S. & Ideker, T. Integrating physical and genetic maps: from genomes to interaction networks. *Nature Rev. Genet.* **8**, 699–710 (2007).
144. Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein networks. *Nature Biotech.* **23**, 561–566 (2005).
145. Ulitsky, I., Shlomi, T., Kupiec, M. & Shamir, R. From E-MAPs to module maps: dissecting quantitative genetic interactions using physical interactions. *Mol. Syst. Biol.* **4**, 209 (2008).
146. 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).
147. Srivas, R. *et al.* Assembling global maps of cellular function through integrative analysis of physical and genetic networks. *Nature Protoc.* **6**, 1308–1323 (2011).
148. 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).
149. Wilmes, G. M. *et al.* A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. *Mol. Cell* **32**, 735–746 (2008).
150. Fiedler, D. *et al.* Functional organization of the *S. cerevisiae* phosphorylation network. *Cell* **136**, 952–963 (2009).
151. Zhang, L. V. *et al.* Motifs, themes and thematic maps of an integrated *Saccharomyces cerevisiae* interaction network. *J. Biol.* **4**, 6 (2005).
152. Yeger-Lotem, E. *et al.* Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proc. Natl Acad. Sci. USA* **101**, 5934–5939 (2004).
153. Tan, K., Shlomi, T., Feizi, H., Ideker, T. & Sharan, R. Transcriptional regulation of protein complexes within and across species. *Proc. Natl Acad. Sci. USA* **104**, 1283–1288 (2007).

154. Herrgard, M. J., Lee, B. S., Portnoy, V. & Palsson, B. O. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.* **16**, 627–635 (2006).
155. Lee, J. M., Gianchandani, E. P., Eddy, J. A. & Papin, J. A. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput. Biol.* **4**, e1000086 (2008).
156. Chandrasekaran, S. & Price, N. D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **107**, 17845–17850 (2010).
157. Deshpande, R., Sharma, S., Verfaillie, C. M., Hu, W. S. & Myers, C. L. A scalable approach for discovering conserved active subnetworks across species. *PLoS Comput. Biol.* **6**, e1001028 (2010).
This study illustrates how a combination of integrative approaches may be simultaneously applied for the identification of conserved active modules.
158. Waltman, P. *et al.* Multi-species integrative biclustering. *Genome Biol.* **11**, R96 (2010).
159. Ryan, C. J. *et al.* Hierarchical modularity and the evolution of genetic interactomes across species. *Mol. Cell* **46**, 691–704 (2012).
160. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006).
161. Ito, T. *et al.* Roles for the two-hybrid system in exploration of the yeast protein interactome. *Mol. Cell. Proteomics.* **1**, 561–566 (2002).
162. Shou, C. *et al.* Measuring the evolutionary rewiring of biological networks. *PLoS Comput. Biol.* **7**, e1001050 (2011).
163. 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).
164. Bebek, G., Koyuturk, M., Price, N. D. & Chance, M. R. Network biology methods integrating biological data for translational science. *Brief Bioinform.* **13**, 446–459 (2012).
165. Ulitsky, I. & Shamir, R. Identification of functional modules using network topology and high-throughput data. *BMC Syst. Biol.* **1**, 8 (2007).
166. Beisser, D., Klau, G. W., Dandekar, T., Müller, T. & Dittrich, M. T. BioNet: an R-Package for the functional analysis of biological networks. *Bioinformatics* **26**, 1129–1130 (2010).
167. Paull, E. O. *et al.* Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE). *Bioinformatics* http://dx.doi.org/10.1093/bioinformatics/btt471 (2013).
168. Kalaev, M., Smoot, M., Ideker, T. & Sharan, R. NetworkBLAST: comparative analysis of protein networks. *Bioinformatics* **24**, 594–596 (2008).
169. Liao, C. S., Lu, K., Baym, M., Singh, R. & Berger, B. IsoRankN: spectral methods for global alignment of multiple protein networks. *Bioinformatics* **25**, i253–i258 (2009).
170. Zhang. *B. et al.* DDN: a caBIG® analytical tool for differential network analysis. *Bioinformatics* **27**,1036–1038 (2011).
171. Gill, R. & Datta, S. A statistical framework for differential network analysis from microarray data. *BMC Bioinformatics* **11**, 95 (2010).

Acknowledgements
We gratefully acknowledge US National Institutes of Health (NIH) grants P41 GM103504 and P50 GM085764 in support of this work.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
The Cancer Genome Atlas (TCGA):
http://cancergenome.nih.gov

SUPPLEMENTARY INFORMATION
See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF