Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Network-based approaches for analysis of complex biological systems

Deborah Chasman[1], Alireza Fotuhi Siahpirani[2,1,3] and Sushmita Roy[1,3,2]

Cells function and respond to changes in their environment by the coordinated activity of their molecular components, including mRNAs, proteins and metabolites. At the heart of proper cellular function are molecular networks connecting these components to process extra-cellular environmental signals and drive dynamic, context-specific cellular responses. Network-based computational approaches aim to systematically integrate measurements from high-throughput experiments to gain a global understanding of cellular function under changing environmental conditions. We provide an overview of recent methodological developments toward solving two major computational problems within this field in the past two years (2013–2015): *network reconstruction* and *network-based interpretation*. Looking forward, we envision development of methods that can predict phenotypes with high accuracy as well as provide biologically plausible mechanistic hypotheses.

Addresses
[1] Wisconsin Institute for Discovery, University of Wisconsin–Madison, Madison, WI 53715, United States
[2] Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI 53706, United States
[3] Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison, Madison, WI 53792, United States

Corresponding author: Roy, Sushmita (sroy@biostat.wisc.edu)

Current Opinion in Biotechnology 2016, **39**:157–166

This review comes from a themed issue on **Systems biology**

Edited by **Fabian Theis** and **Mark Styczynski**

http://dx.doi.org/10.1016/j.copbio.2016.04.007

Introduction

Normal and disease cellular states are the integrated outputs of networks acting at multiple levels of regulation, including the pre-transcriptional, transcriptional and post-translational levels [1–4]. Advances in omic techniques are providing an unprecedented capacity to measure RNA (coding and non-coding), protein, and post-translational modification levels under different biological contexts such as time, cell states, tissues, or organisms [5–8]. Systematic integration of such datasets is essential to identify molecular networks controlling normal and disease states, and, ultimately, predict complex phenotypes from molecular markers [9,10]. This review focuses on recent efforts (2013–2015, Supplementary Table 1) in the field of network biology to address two major challenges that have emerged in the era of high-throughput biology (Figure 1): Network reconstruction, which entails inference of shared and context-specific network connectivity from multiple omic measurements of molecular entities, and Network-based interpretation of experimental observations to guide further study.

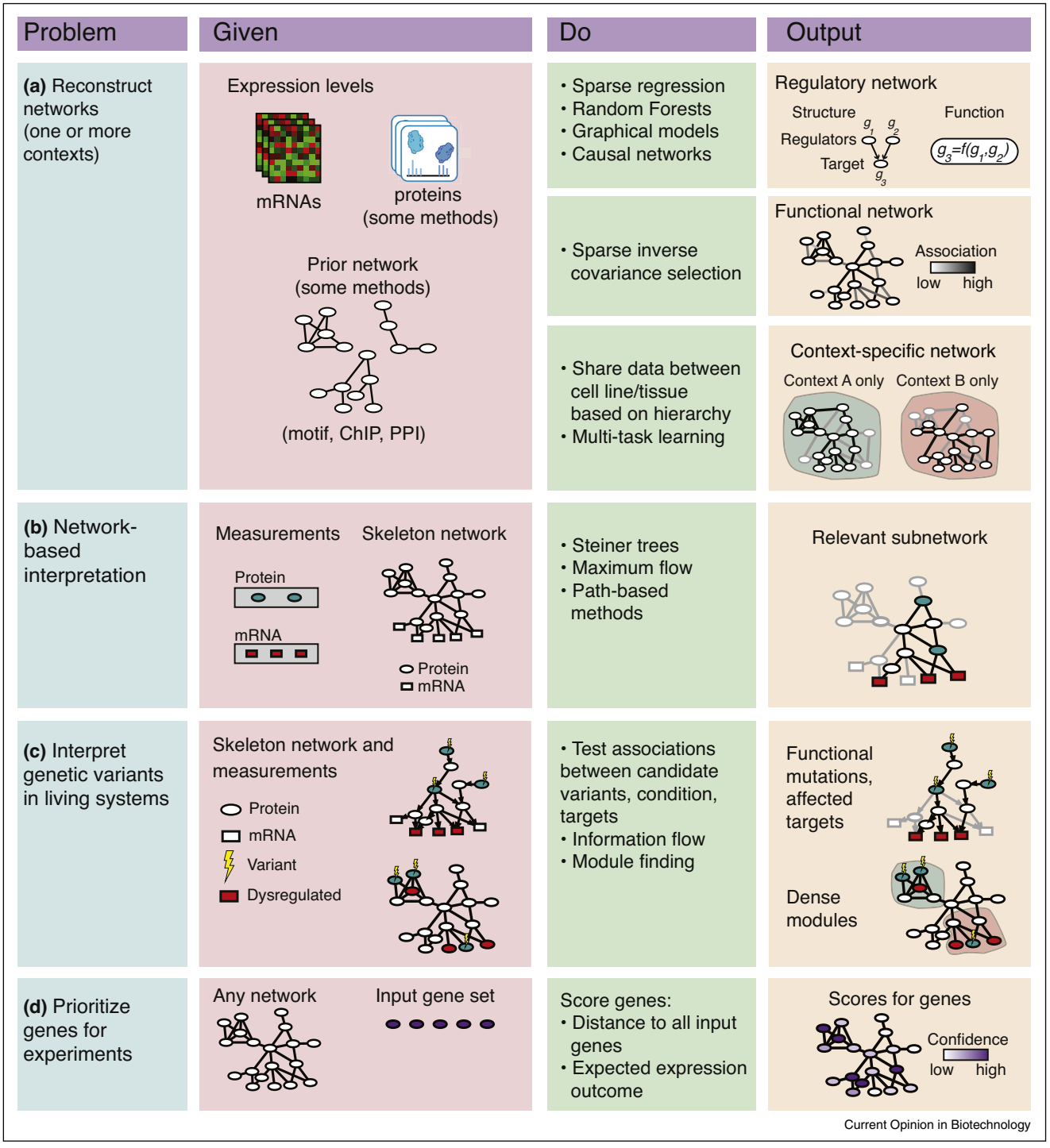
Network reconstruction

A long-standing problem in systems biology is to infer causal, regulatory connections among genes, proteins and metabolites and to understand how connections change over time, cell types and other conditions. A network provides a natural representation of a complex cellular system with nodes representing the molecular components and edges representing different types of connectivities (Box 1a). The goal of *network reconstruction* (also referred to as *network inference*) methods is to infer such connections from genome-scale measurements of genes, proteins and metabolites from multiple conditions (Figure 1a). In this review, we focus on transcriptional networks and briefly discuss extensions to post-transcriptional, signaling and metabolic networks. We first discuss recent work in the widely studied problem of reverse engineering a single genome-wide network from large collections of gene expression data, followed by approaches to infer integrative and context-specific networks, problems of emerging interest.

Regulatory network reconstruction for one condition

Since the advent of microarrays and more recently RNA-seq to measure genome-wide transcriptomes, *expression-based network inference* methods have been popular for inferring *regulatory networks* among regulatory proteins such as transcription factors (TFs) and signaling proteins and target genes. A regulatory network model has two components (Box 1b): the structure, which specifies the regulators of a target gene, and the regulatory function, encoded as a mathematical function, which describes how individual and combined regulatory inputs specify a target gene's expression. Several different mathematical functions have been proposed to relate regulatory inputs to expression output including Boolean functions,

Figure 1



Overview of major computational network biology tasks addressed here. Shown are the key inputs, algorithmic approaches and outputs produced in each type of task. **(a)** Network reconstruction for identifying a single network or multiple context-specific networks. **(b)** Network-based interpretation to identify a relevant subnetwork that integrates diverse data types (e.g. mRNA and proteomic measurements) on a skeleton network. **(c)** Network-based interpretation of genetic variants to identify either subnetworks or network modules. **(d)** Network-based prioritization to prioritize genes for further experiments based on an input gene set and measured expression levels and proximity of candidate genes on the skeleton graph.

Box 1 Network terminology

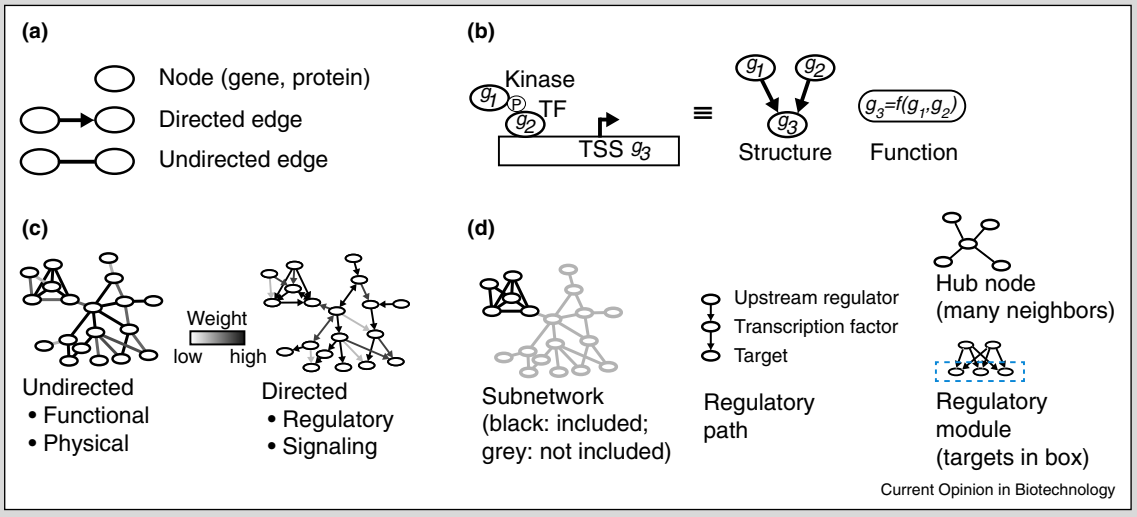
A molecular network is defined by the type of entities it connects and the information it contains about those connections.

a. Network elements. Nodes in a molecular network are intracellular molecules, frequently genes or proteins. Edges represent interactions (binding, regulatory, post-translational modification, correlation) and may be directed (if it is known that the first molecule's activity or function influences the second's) or undirected.

b. Network structure and function. On the right is a cartoon showing a kinase g_1 and transcription factor (TF) g_2 whose coordinate activity regulates a gene g_3. The structure shows g_1 and g_2 as regulators of g_3, and the function f quantifies the relationship between regulator state and target. This function estimates the state or activity level of the target gene in one or multiple time points/conditions as a function of state/activity of regulators, or other input parameters. In the literature, this function has taken many forms, including simple linear regression, linear or non-linear differential equation, and probabilistic Boolean function.

c. Types of networks. Regulatory networks describe transcriptional regulatory programs (as in Panel b). Signaling networks describe cascades of post-translational modifications from signaling receptors to transcription factors. Functional networks are undirected, and represent statistical dependencies between nodes. Physical networks represent experimentally observed binding interactions between proteins or between proteins and DNA or RNA. In any type of network, weights on edges may correspond to statistical significance, confidence, or inferred importance of the edge.

d. Organizational units. Networks are composed of smaller structural units. A subnetwork is a subset of nodes and edges from the original network; subnetworks are used to predict functional coordination between nodes. A regulatory path is a chain of nodes and edges that posits the ordered mechanism by which upstream regulators and transcription factors influence the regulation of a target gene. A regulatory module is a group of genes that are co-expressed and are predicted to share the same regulatory program. A hub node has a high degree (number of neighbors); hubs are often predicted to be important regulators of one or more cellular pathways.



ordinary differential equations, thermodynamic models, and probabilistic functions (reviewed by Kim *et al.* [11]).

Expression-based network inference has been approached in two ways: by reconstructing regulatory programs for (i) individual genes at a time ('per-gene' methods) and (ii) modules of genes ('per-module') (reviewed by DeSmet and Marchal [12]). While per-gene methods provide fine-grain gene-level regulatory information, per-module methods have the benefit of being easier to interpret and capture modular organization of networks (Box 1d). Some recent work combines the two paradigms [13], providing a single method to infer regulatory programs for individual genes and gene modules. Although the idea of inferring networks from expression alone is attractive, genome-scale inference of such networks using current methods is an open challenge. In particular, the DREAM5 competition, which performed one of the largest comparative studies to date of network inference methods, showed that the agreement between the inferred networks and experimentally measured TF–target gene interactions is close to random in unicellular eukaryotes such as yeast [14].

One direction of work to address this issue is to impose additional constraints in the form of *prior knowledge* on the network structure to favor certain edges with auxiliary support from other data types, such as the presence of DNA-binding motifs for specific transcription factors. While the idea of incorporating priors into a probabilistic

model, such as a Bayesian network [15], was proposed several years ago [16,17], only recently has it been used on a genome-wide scale to integrate diverse types of evidence. Recent studies have implemented prior usage within a dependency network learning framework [18], where the network structure is inferred by solving a set of linear or non-linear regression functions: Greenfield *et al.*'s [19] extension of regularized linear regression in Modified Elastic Net (MEN) and Bayesian Best Subset Regression (BBSR), and iRafNet [20] (extending GENIE3 [21]). iRafNet incorporates different types of prior networks (protein–protein interactions, knockout data, gene expression time courses), whereas MEN and BBSR take a single input prior network.

Prior-based methods still rely on the predictive power of the mRNA level of the transcription factor, which might not reflect the cellular activity of a TF [22]. Towards this end, Arrieta-Ortiz and Hafemeister *et al.* [23] first use network component analysis [24] to predict the activity of TFs given a prior network, and then use these TF activities to predict a regulatory network for the bacteria *Bacillus subtilis*. A large fraction (~62%) of the predicted interactions had experimental support indicating high accuracy in their predictions. While this shows early promise for prior-based methods for bacteria, an important future direction is to apply such approaches to complex eukaryotes (e.g. in plants and mammals). A key requirement of such approaches is prior knowledge of *cis*-regulatory elements on DNA sequence specificity of TFs. Experimental approaches to measure sequence specificity of TFs *in vitro* (e.g. using protein binding microarrays [25]) and *in vivo* (e.g. open chromatin using DNase I [26], ATAC-seq [27] and chromatin state [28]), combined with computational modeling [29,30], could significantly improve the quality of the prior networks for transcriptional regulatory networks.

More complete and accurate networks require measuring and modeling additional levels of transcriptional and post-transcriptional regulation. Accordingly, new methods are being developed to integrate these aspects. For example, using genome-wide miRNA and mRNA levels measurements from cancer samples, several groups have inferred miRNA-target gene networks [31,32]. Osmanbeyoglu *et al.* [33•] infer latent activity of and relationships between signaling proteins and transcription factors using a new regularized regression approach, integrating both reverse phase protein array levels and gene expression. Metabolic networks also have regulatory roles and interact with transcriptional networks to control the overall state of organisms [34,35]. Recent availability of large scale metabolomic profiles has fueled parallel development of network-based methods to analyze metabolomic data. We refer the reader to recent reviews on network approaches with metabolomic data [36,37] and integration of these networks with transcriptional networks [38,39]. A direction that can be particularly useful for capturing different levels of regulation is to better model the distribution of the random variables representing the molecular entities, for example, by allowing heterogeneous random variables [40,41].

Inference of a causal network by asserting directionality on edges is another important challenge in network reconstruction. Gene expression data from natural variation populations [42] or single gene perturbations [43] can help assert directed causal connections. To this end, the caPC (covariate-adjusted PC) algorithm [44] infers a partially directed graph by regressing each gene's expression on a set of single-nucleotide polymorphisms (SNPs), and then uses the PC algorithm [45] to infer directionality based on conditional independence tests. Temporal information can also capture causal dependencies and dynamic Bayesian frameworks are particularly promising to address this [46•]. An alternative is to focus on inferring the directed connections among regulators only, using a prior regulatory network as input and inferring the activity of regulators rather than inferring new edges to targets. The biRte method [47] extends Nested Effects Models [48], to infer a directed graph over the regulators based on the identity and expression of their targets. An advantage of this method is that it does not predict regulator activity only based on mRNA levels. These approaches are promising steps toward solving a key challenge of inferring directionality in regulatory networks.

Context-specificity and dynamics in network reconstruction

The previous section focused on learning a single network from data. Often it is important to infer how networks change between different contexts (e.g. cell types and diseases) or over time. Recently, several methods have been developed to jointly infer networks that each represents a different context. Because information is shared during the inference process, the networks can be used to study shared and context-specific components.

The vast majority of methods have used Gaussian Graphical models (GGMs) [49,50] that infer undirected 'functional' interactions corresponding to direct statistical dependencies (Box 1c). Learning a GGM translates to estimating the non-zero off-diagonal entries in the inverse of the covariance matrix (also called the precision matrix), further imposing regularization terms to encourage sparsity. One approach by Kling *et al.* [51•] uses an augmented Sparse Inverse Covariance Selection (SICS) approach to jointly estimate one network (precision matrix) for each cancer, each network predicting connections between genes, microRNAs and sequence mutations. This is accomplished by a regularized regression framework with a novel multi-graph prior that encourages similarity between the networks as well as a modular structure within each network. Other approaches incorporate prior

information about the relationship between the contexts. Treegl [52] and GNAT [53•] use a known hierarchy over contexts, such as a cell lineage. In Treegl, data are observed at both internal and leaf nodes, while for GNAT, data are available only in the leaf nodes. Aiming to get closer to a directed regulatory network, Ontogenet [54•] infers per-module regulatory programs for multiple cell types in a lineage. This approach is similar to Treegl and GNAT, except that in Ontogenet the covariates of a gene are restricted to annotated regulators such as transcription factors and one estimates regulators for modules rather than individual genes, making this approach amenable to small sample size problems. GNAT and Ontogenet are of particular interest because they demonstrate that these approaches can be used on a large scale.

Context-specific networks have also been examined using supervised learning by training on a gold-standard network extracted from known interaction databases (e.g. [55,56]) that is filtered based on node expression in a specific condition [57]. Park *et al.* [58] train tissue-specific support vector machines to predict the probability of an edge belonging to a particular interaction type (transcriptional regulation, phosphorylation, protein co-complexes, and post-translational regulation). Greene *et al.* [59•] apply a Bayesian data integration approach to learn tissue-specific networks for 144 tissues based on available human gene expression data sets in the Gene Expression Omnibus [60].

A more fine-grained notion of context is time; however, the above approaches are not applicable here because of the lack of sufficient samples per time point. Yosef *et al.* [61] overlayed time-point specific gene expression on a skeleton regulatory network inferred from sequence-specific motifs and targets. ODE-based models have the finest possible resolution [62], however, learning such models on a genome-scale is intractable and most approaches have focused on few dozen nodes [46•]. To overcome the computational complexity while modeling different types of dependencies including temporal dependencies, the Jump3 method [63] uses tree-based models on time-series data and was able to model hundreds of genes.

Of special interest are methods that use single cell data such as single cell RNA-seq [64,65] or proteomic levels [66] for inferring regulatory networks in cell fate specification problems [66,67••,68]. The method of Ocone *et al.* [67••] uses dimensionality reduction techniques and cellular trajectory learning [69] to infer a pseudo time course. They then use an ODE-based model to infer a regulatory network for each of the developmental branches. Because these assays can measure hundreds to thousands of cells, each cell providing a simultaneous measurement of multiple molecules at a time, these technologies open up a new avenue of inferring networks per cell type.

In summary, network reconstruction is the problem of inferring regulatory causal connections between molecular nodes in one or multiple conditions. Expression-based network inference methods have been popular because mRNA levels are the most widely available type of data. However, expression alone is not sufficient to learn regulatory networks, and new approaches that either incorporate prior interactions or additional types of data will be important to learn more accurate and complete networks.

Network-based interpretation and prioritization

Once a network structure is available, powerful computational approaches have been developed to tackle a variety of problems, broadly grouped under 'network-based interpretation and prioritization'. Here we address three broad classes of methods: integrating gene sets from multiple high-throughput experiments, examining perturbation in networks, and prioritizing genes for follow up experiments. Common to these approaches is to use the network as a 'skeleton' to define the universe of possible physical (e.g. protein–protein, protein–DNA) and functional (genetic) relationships among genes (Box 1c).

Integrating gene hits from complementary high-throughput experiments

Frequently, multiple high-throughput experimental assays such as transcriptomic or proteomic profiling, as well as large-scale functional screening studies [70,71], are used to identify genes important for a specific biological process [72]. Often, the experimental methodologies are complementary and identify only partially overlapping *hit* gene sets. An important question is if, and how, the genes identified from each assay are related to each other. *Network flow* methods (Figure 1d) are popular approaches to identify subnetworks integrating input gene sets based on their connectivity on a skeleton network (reviewed by Kim *et al.* [73]). The key idea is that the subnetwork depicts how information flows between the different hit sets. For example, a subnetwork can link functional screen hits (upstream) to differential mRNA expression (downstream). Network flow methods often (but not exclusively) seek to find a minimal or sparse subnetwork (i.e. using a small number of edges to connect many hits) and vary in the node-level, edge-level and path-level properties that are inferred for each subnetwork. For example, prize-collecting Steiner trees [74] and maximum flow [72] methods define node-level or edge-level properties (e.g. sign, directionality) and define an objective function that prefers a minimal network. In contrast, candidate path-based methods can incorporate more complex path-level properties [75–79], for example, enabling the connection of specific node pairs. However, enumerating candidate paths or optimizing over them can be computationally challenging. As in network reconstruction, an emerging theme among information flow methods is to extend existing approaches to jointly infer subnetworks for multiple related contexts [79–81].

All methods define these network level properties as constraints and solve the constrained optimization method using either: linear or integer linear programming (ILP) methods [72,74,76–79], or statistical inference within a maximum likelihood framework [75,82,83]. When considering genome-scale networks, ILP-based methods have the advantage in that they can find globally optimal solutions more efficiently compared to probabilistic approaches. However, probabilistic approaches are also useful at smaller scales as different parameters can be estimated in a data-driven manner.

Interpreting genetic variation

A network also provides molecular context to interpret genetic variation (Figure 1d), including Single Nucleotide Polymorphisms (SNPs). Some approaches are applicable when there is a known directionality from the perturbation to specific affected targets. PINE (*P*erturbations in *NE*tworks) [84] is a probabilistic method to predict which branches of an input parameterized signaling network are affected by genetic variants under specific conditions. If a less detailed input network is sufficient, network flow methods (e.g. *minimal set cover*) can also be used. For example, in cancer, identifying causal 'driver' variants from correlated 'passenger' mutations is important. This problem is commonly addressed by *minimum set cover* approach (recently, [85,86]) to distinguish causal driver mutations from passenger mutations by finding a small set of drivers that can be connected to the targets by paths in a protein–protein network.

If only a list of mutations is available, or if causal directionality from mutations to gene expression is not assumed, one can use *module-based methods* to find densely connected components of the network that are enriched for highly mutated genes (e.g. [87,88]). Grouping the mutated or dysregulated genes into modules may make the approach more resistant to noise in the network or data. Recently this concept has been applied in conjunction with expression data to interpret genetic variation in autism [89] and glioblastoma [90].

A limitation of current skeleton-network based approaches is that they typically include only perturbations to coding regions, and cannot be used to interpret regulatory variants, many of which can lie far away from genes [91]. As our ability to identify regulatory variation improves, for example from high-throughput regulatory genomic datasets [5,6], extending network-based methods to incorporate regulatory variants will become increasingly important [92,93].

Gene prioritization

Protein–protein and functional interaction networks are often used to prioritize genes for further experiments (Figure 1e). Most of these approaches (reviewed by Moreau and Tranchevant [94]) have either relied on network centrality measures (e.g. connectivity of hub genes, Box 1d, [95]), or, have used an input set of known relevant genes to rank candidates using the concept of *guilt by association*: genes that are close to input genes (e.g. based on shortest path length, or more global diffusion-based measures [96]) in the skeleton network are likely to have similar function. An increasingly important and exciting new direction is to prioritize genes to efficiently convert one cell type into another, for example, to reprogram differentiated cells into induced pluripotent cells [97,98••]. CellNet [98••] first infers tissue and cell type-specific networks, and then uses those networks in a Random Forest classification approach to prioritize genes for reprogramming and to quantify similarity between a query sample and a known cell type. A subsequent approach, Mogrify [97], prioritizes genes based on proximity on an input skeleton network and a measure of differential expression, which is informed by the lineage structure of cell types. Both approaches successfully recapitulated known cell-fate specification genes, and also predicted new genes that were validated to be important regulators for establishing a specific fate.

In summary, network-based methods can be powerful for interpretation of high-throughput measurements and for prioritization of experiments. The choice of the skeleton network is important as most approaches rely on interactions from public databases that may not be relevant to the cellular context being modeled. Recent efforts to infer context-specific interactions may allow these network-based approaches to achieve higher precision [59•].

Conclusions

Network-based computational methods aim to build, interpret and integrate molecular networks and have been important tools for studies of complex responses. In this review, we discussed recent promising approaches for two major problems: network reconstruction and network-based interpretation. Going forward we envision addressing one major challenge: to build network-based causal, predictive models of complex phenotypes. This will require us to construct networks and prioritize experiments in concert, rather than independently, within an integrated computational-experimental framework that permits multiple iterations of prediction and validation. Novel network-based models that leverage new types of datasets from emerging technologies for measuring RNA [65] and open chromatin [99] for single cells, as well as approaches to perform targeted perturbations (e.g. using CRISPR/Cas9 technologies [100,101]) will increasingly enable us to infer causal edges and identify the mechanistic, molecular underpinnings of complex traits and diseases.

Acknowledgements

SR is supported by a National Science Foundation CAREER award (NSF DBI: 1350677) and a Sloan Foundation research fellowship. This publication was made possible in part by US Environmental Protection Agency grant 83573701.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.copbio.2016.04.007.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Barabasi AL, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach to human disease**. *Nat Rev Genet* 2011, **12**:56-68 http://dx.doi.org/10.1038/nrg2918.

2. Bensimon A, Heck AJR, Aebersold R: **Mass spectrometry-based proteomics and network biology**. *Annu Rev Biochem* 2012, **81**:379-405 http://dx.doi.org/10.1146/annurev-biochem-072909-100424.

3. Ideker T, Krogan NJ: **Differential network biology**. *Mol Syst Biol* 2012, **8**:565 http://dx.doi.org/10.1038/msb.2011.99.

4. Lee TI, Young RA: **Transcriptional regulation and its misregulation in disease**. *Cell* 2013, **152**:1237-1251 http://dx.doi.org/10.1016/j.cell.2013.02.014.

5. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M: **Integrative analysis of 111 reference human epigenomes**. *Nature* 2015, **518**:317-330 http://dx.doi.org/10.1038/nature14248.

6. ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome**. *Nature* 2012, **489**:57-74 http://dx.doi.org/10.1038/nature11247.

7. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM: **The Cancer Genome Atlas Pan-Cancer analysis project**. *Nat Genet* 2013, **45**:1113-1120 http://dx.doi.org/10.1038/ng.2764.

8. GTEx Consortium: **The Genotype-Tissue Expression (GTEx) project**. *Nat Genet* 2013, **45**:580-585 http://dx.doi.org/10.1038/ng.2653.

9. Cho DY, Kim YA, Przytycka TM: **Chapter 5: network biology approach to complex diseases**. *PLoS Comput Biol* 2012, **8**:e1002820 http://dx.doi.org/10.1371/journal.pcbi.1002820.

10. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D: **Methods of integrating data to uncover genotype–phenotype interactions**. *Nat Rev Genet* 2015, **16**:85-97 http://view.ncbi.nlm.nih.gov/pubmed/25582081.

11. Kim HD, Shay T, O'Shea EK, Regev A: **Transcriptional regulatory circuits: predicting numbers from alphabets**. *Science* 2009, **325**:429-432 http://dx.doi.org/10.1126/science.1171347.

12. De Smet R, Marchal K: **Advantages and limitations of current network inference methods**. *Nat Rev Microbiol* 2010, **8**:717-729 http://dx.doi.org/10.1038/nrmicro2419.

13. Roy S, Lagree S, Hou Z, Thomson JA, Stewart R, Gasch AP: **Integrated module and gene-specific regulatory inference implicates upstream signaling networks**. *PLoS Comput Biol* 2013, **9**:e1003252 http://dx.doi.org/10.1371/journal.pcbi.1003252.

14. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, DREAM5 Consortium, Kellis M, Collins JJ, Stolovitzky G: **Wisdom of crowds for robust gene network inference**. *Nat Methods* 2012, **9**:796-804 http://dx.doi.org/10.1038/nmeth.2016.

15. Friedman N: **Inferring cellular networks using probabilistic graphical models**. *Science* 2004, **303**:799-805 http://dx.doi.org/10.1126/science.1094068.

16. Werhli AV, Husmeier D: **Reconstructing gene regulatory networks with Bayesian networks by combining expression data with multiple sources of prior knowledge**. *Stat Appl Genet Mol Biol* 2007, **6** http://dx.doi.org/10.2202/1544-6115.1282 Article 15.

17. Mukherjee S, Speed TP: **Network inference using informative priors**. *Proc Natl Acad Sci U S A* 2008, **105**:14313-14318 http://dx.doi.org/10.1073/pnas.0802272105.

18. Heckerman D, Chickering DM, Meek C, Rounthwaite R, Kadie C: **Dependency networks for collaborative filtering and data visualization**. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc; 2000:264-273.

19. Greenfield A, Hafemeister C, Bonneau R: **Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks**. *Bioinformatics* 2013, **29**:1060-1067 http://dx.doi.org/10.1093/bioinformatics/btt099.

20. Petralia F, Wang P, Yang J, Tu Z: **Integrative random forest for gene regulatory network inference**. *Bioinformatics* 2015, **31**:i197-i205 http://dx.doi.org/10.1093/bioinformatics/btv268.

21. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P: **Inferring regulatory networks from expression data using tree-based methods**. *PLoS ONE* 2010, **5** http://dx.doi.org/10.1371/journal.pone.0012776.

22. Vogel C, Marcotte EM: **Insights into the regulation of protein abundance from proteomic and transcriptomic analyses**. *Nat Rev Genet* 2012, **13**:227-232 http://dx.doi.org/10.1038/nrg3185.

23. Arrieta-Ortiz ML, Hafemeister C, Bate AR, Chu T, Greenfield A, Shuster B, Barry SN, Gallitto M, Liu B, Kacmarczyk T *et al.*: **An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network**. *Mol Syst Biol* 2015, **11**:839 http://dx.doi.org/10.15252/msb.20156236.

24. Liao JC, Boscolo R, Yang YL, Tran LM, Sabatti C, Roychowdhury VP: **Network component analysis: reconstruction of regulatory signals in biological systems**. *Proc Natl Acad Sci U S A* 2003, **100**:15522-15527 http://dx.doi.org/10.1073/pnas.2136632100.

25. Siggers T, Gordân R: **Protein–DNA binding: complexities and multi-protein codes**. *Nucleic Acids Res* 2014, **42**:2099-2111 http://dx.doi.org/10.1093/nar/gkt1112.

26. Vierstra J, Stamatoyannopoulos JA: **Genomic footprinting**. *Nat Methods* 2016, **13**:213-221 http://dx.doi.org/10.1038/nmeth.3768.

27. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position**. *Nat Methods* 2013, **10**:1213-1218 http://dx.doi.org/10.1038/nmeth.2688.

28. Mendenhall EM, Bernstein BE: **Chromatin state maps: new technologies, new insights**. *Curr Opin Genet Dev* 2008, **18**:109-115 http://dx.doi.org/10.1016/j.gde.2008.01.010.

29. Gusmao EG, Allhoff M, Zenke M, Costa IG: **Analysis of computational footprinting methods for DNase sequencing experiments**. *Nat Methods* 2016, **13**:303-309 http://dx.doi.org/10.1038/nmeth.3772.

30. González AJ, Setty M, Leslie CS: **Early enhancer establishment and regulatory locus complexity shape transcriptional**

programs in hematopoietic differentiation. *Nat Genet* 2015, **47**:1249-1259 http://dx.doi.org/10.1038/ng.3402.

31. Lee E, Ito K, Zhao Y, Schadt EE, Irie HY, Zhu J: **Inferred miRNA activity identifies miRNA-mediated regulatory networks underlying multiple cancers**. *Bioinformatics* 2015 http://dx.doi.org/10.1093/bioinformatics/btv531.

32. Wang Z, Xu W, Liu Y: **Integrating full spectrum of sequence features into predicting functional microRNA–mRNA interactions**. *Bioinformatics* 2015 http://dx.doi.org/10.1093/bioinformatics/btv392.

33. • Osmanbeyoglu HU, Pelossof R, Bromberg JF, Leslie CS: **Linking signaling pathways to transcriptional programs in breast cancer**. *Genome Res* 2014, **24**:1869-1880 http://dx.doi.org/10.1101/gr.173039.114.
A novel regularized regression framework (affinity regression) for inferring relationships between signaling proteins and transcription factors using protein and mRNA levels. Notably, this approach can make predictions for both types of data. Given a new protein (or gene) expression profile, it can be used to predict active transcription factors (or signaling proteins).

34. Watson E, MacNeil LT, Arda HE, Zhu LJ, Walhout AJM: **Integration of metabolic and gene regulatory networks modulates the *C. elegans* dietary response**. *Cell* 2013, **153**:253-266 http://dx.doi.org/10.1016/j.cell.2013.02.050.

35. Oliveira AP, Dimopoulos S, Busetto AG, Christen S, Dechant R, Falter L, Haghir Chehreghani M, Jozefczuk S, Ludwig C, Rudroff F, Schulz JC, González A, Soulard A, Stracka D, Aebersold R, Buhmann JM, Hall MN, Peter M, Sauer U, Stelling J: **Inferring causal metabolic signals that regulate the dynamic TORC1-dependent transcriptome**. *Mol Syst Biol* 2015, **11**:802.

36. O'Brien EJ, Monk JM, Palsson BO: **Using genome-scale models to predict biological capabilities**. *Cell* 2015, **161**:971-987 http://dx.doi.org/10.1016/j.cell.2015.05.019.

37. King ZA, Lloyd CJ, Feist AM, Palsson BO: **Next-generation genome-scale models for metabolic engineering**. *Curr Opin Biotechnol* 2015, **35**:23-29 http://dx.doi.org/10.1016/j.copbio.2014.12.016.

38. Gonçalves E, Bucher J, Ryll A, Niklas J, Mauch K, Klamt S, Rocha M, Saez-Rodriguez J: **Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models**. *Mol Biosyst* 2013, **9**:1576-1583.

39. Kim J, Reed JL: **Refining metabolic models and accounting for regulatory effects**. *Curr Opin Biotechnol* 2014, **29**:34-38.

40. Chen S, Witten DM, Shojaie A: **Selection and estimation for mixed graphical models**. *Biometrika* 2014. asu051.

41. Žitnik M, Zupan B: **Gene network inference by fusing data from diverse distributions**. *Bioinformatics* 2015, **31**:i230-i239 http://dx.doi.org/10.1093/bioinformatics/btv258.

42. Brem RB, Kruglyak L: **The landscape of genetic complexity across 5,700 gene expression traits in yeast**. *Proc Natl Acad Sci U S A* 2005, **102**:1572-1577.

43. Kemmeren P, Sameith K, van de Pasch LA, Benschop JJ, Lenstra TL, Margaritis T, ODuibhir E, Apweiler E, van Wageningen S, Ko CW *et al.*: **Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors**. *Cell* 2014, **157**:740-752.

44. Gao B, Cui Y: **Learning directed acyclic graphical structures with genetical genomics data**. *Bioinformatics* 2015 http://dx.doi.org/10.1093/bioinformatics/btv513.

45. Spirtes P, Glymour CN, Scheines R: **Causation, Prediction, and Search**. MIT Press; 2000.

46. • Oates CJ, Dondelinger F, Bayani N, Korkola J, Gray JW, Mukherjee S: **Causal network inference using biochemical kinetics**. *Bioinformatics* 2014, **30**:i468-i474 http://dx.doi.org/10.1093/bioinformatics/btu452.
Bayesian framework for inference of causal networks with non-linear, biochemical kinetic dependencies.

47. Fröhlich H: **biRte: Bayesian inference of context-specific regulator activities and transcriptional networks**. *Bioinformatics* 2015 http://dx.doi.org/10.1093/bioinformatics/btv379.

48. Markowetz F, Kostka D, Troyanskaya OG, Spang R: **Nested effects models for high-dimensional phenotyping screens**. *Bioinformatics* 2007, **23**:i305-i312.

49. Yuan M, Lin Y: **Model selection and estimation in the Gaussian graphical model**. *Biometrika* 2007, **94**:19-35.

50. Friedman J, Hastie T, Tibshirani R: **Sparse inverse covariance estimation with the graphical lasso**. *Biostatistics* 2008, **9**:432-441.

51. • Kling T, Johansson P, Sanchez J, Marinescu VD, Jörnsten R, Nelander S: **Efficient exploration of pan-cancer networks by generalized covariance selection and interactive web content**. *Nucleic Acids Res* 2015, **43**:e98 http://dx.doi.org/10.1093/nar/gkv413.
Method for integrating multiple genomic measurements across many sample cohorts by jointly estimating regularized inverse covariance matrices.

52. Parikh AP, Curtis RE, Kuhn I, Becker-Weimann S, Bissell M, Xing EP, Wu W: **Network analysis of breast cancer progression and reversal using a Tree-Evolving network algorithm**. *PLoS Comput Biol* 2014, **10**:e1003713 http://dx.doi.org/10.1371/journal.pcbi.1003713.

53. • Pierson E, GTEX Consortium, Koller D, Battle A, Mostafavi S, Ardlie KG, Getz G, Wright FA, Kellis M, Volpi S, Dermitzakis ET: **Sharing and specificity of co-expression networks across 35 human tissues**. *PLoS Comput Biol* 2015, **11**:e1004220 http://dx.doi.org/10.1371/journal.pcbi.1004220.
Transfer-learning approach to identify context-specific co-expression networks given a hierarchy over contexts. Application to 35 GTEx tissues identified tissue-specific components that could not be explained due to differences in nodes.

54. • Jojic V, Shay T, Sylvia K, Zuk O, Sun X, Kang J, Regev A, Koller D, Immunological Genome Project Consortium, Best AJ, Knell J, Goldrath A, Joic V, Koller D, Shay T, Regev A, Cohen N, Brennan P, Brenner M, Kim F, Rao TN, Wagers A, Heng T, Ericson J, Rothamel K, Ortiz-Lopez A, Mathis D, Benoist C, Bezman NA, Sun JC, Min-Oo G, Kim CC, Lanier LL, Miller J, Brown B, Merad M, Gautier EL, Jakubzick C, Randolph GJ, Monach P, Blair DA, Dustin ML, Shinton SA, Hardy RR, Laidlaw D, Collins J, Gazit R, Rossi DJ, Malhotra N, Sylvia K, Kang J, Kreslavsky T, Fletcher A, Elpek K, Bellemarte-Pelletier A, Malhotra D, Turley S: **Identification of transcriptional regulators in the mouse immune system**. *Nat Immunol* 2013, **14**:633-643 http://dx.doi.org/10.1038/ni.2587.
A multi-task learning approach for inferring modular regulatory programs for 195 cell types in a developmental lineage. Importantly, the approach requires only a few samples per cell type.

55. Chatr-aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L *et al.*: **The BioGRID interaction database: 2015 update**. *Nucleic Acids Res* 2014. gku1204.

56. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M: **KEGG as a reference resource for gene and protein annotation**. *Nucleic Acids Res* 2015 http://dx.doi.org/10.1093/nar/gkv1070.

57. Zhu F, Shi L, Engel JD, Guan Y: **Regulatory network inferred using expression data of small sample size: application and validation in erythroid system**. *Bioinformatics* 2015, **31**:2537-2544 http://dx.doi.org/10.1093/bioinformatics/btv186.

58. Park CY, Krishnan A, Zhu Q, Wong AK, Lee YS, Troyanskaya OG: **Tissue-aware data integration approach for the inference of pathway interactions in metazoan organisms**. *Bioinformatics* 2015, **31**:1093-1101 http://dx.doi.org/10.1093/bioinformatics/btu786.

59. • Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, Zhang R, Hartmann BM, Zaslavsky E, Sealfon SC, Chasman DI, FitzGerald GA, Dolinski K, Grosser T, Troyanskaya OG: **Understanding multicellular function and disease with human tissue-specific networks**. *Nat Genet* 2015, **47**:569-576 http://dx.doi.org/10.1038/ng.3259.
Large-scale prediction of tissue-specific interaction networks using massive amounts of publicly available microarray data and a network-based prioritization tool for sequence variants. The authors provide a web interface (GIANT) for their networks and prediction tools.

60. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository**. *Nucleic Acids Res* 2002, **30**:207-210.

61. Yosef N, Shalek AK, Gaublomme JT, Jin H, Lee Y, Awasthi A, Wu C, Karwacz K, Xiao S, Jorgolli M, Gennert D, Satija R, Shakya A, Lu DY, Trombetta JJ, Pillai MR, Ratcliffe PJ, Coleman ML, Bix M, Tantin D, Park H, Kuchroo VK, Regev A: **Dynamic regulatory network controlling TH17 cell differentiation**. *Nature* 2013, **496**:461-468 http://dx.doi.org/10.1038/nature11981.

62. de Jong H: **Modeling and simulation of genetic regulatory systems: a literature review**. *J Comput Biol* 2002, **9**:67-103 http://dx.doi.org/10.1089/10665270252833208.

63. Huynh-Thu VA, Sanguinetti G: **Combining tree-based and dynamical systems for the inference of gene regulatory networks**. *Bioinformatics* 2015, **31**:1614-1622 http://dx.doi.org/10.1093/bioinformatics/btu863.

64. Göttgens B: **Regulatory network control of blood stem cells**. *Blood* 2015, **125**:2614-2620 http://dx.doi.org/10.1182/blood-2014-08-570226.

65. Stegle O, Teichmann SA, Marioni JC: **Computational and analytical challenges in single-cell transcriptomics**. *Nat Rev Genet* 2015, **16**:133-145 http://dx.doi.org/10.1038/nrg3833.

66. Krishnaswamy S, Spitzer MH, Mingueneau M, Bendall SC, Litvin O, Stone E, Pe'er D, Nolan GP: **Systems biology. Conditional density-based analysis of T cell signaling in single-cell data**. *Science* 2014, **346**:1250689 http://dx.doi.org/10.1126/science.1250689.

67. •• Ocone A, Haghverdi L, Mueller NS, Theis FJ: **Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data**. *Bioinformatics* 2015, **31**:i89-i96 http://dx.doi.org/10.1093/bioinformatics/btv257.
Using the idea that individual cells in a single sample are at different stages in a developmental or response process, the authors present an approach for learning a dynamic regulatory network from single-cell RNA-seq data.

68. Chen H, Guo J, Mishra SK, Robson P, Niranjan M, Zheng J: **Single-cell transcriptional analysis to uncover regulatory circuits driving cell fate decisions in early mouse development**. *Bioinformatics* 2015, **31**:1060-1066 http://dx.doi.org/10.1093/bioinformatics/btu777.

69. Bendall SC, Davis KL, Amir EaD, Tadmor MD, Simonds EF, Chen TJ, Shenfeld DK, Nolan GP, Peer D: **Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development**. *Cell* 2014, **157**:714-725.

70. Giaever G, Nislow C: **The yeast deletion collection: a decade of functional genomics**. *Genetics* 2014, **197**:451-465 http://dx.doi.org/10.1534/genetics.114.161620.

71. Schmidt EE, Pelz O, Buhlmann S, Kerr G, Horn T, Boutros M: **GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update**. *Nucleic Acids Res* 2013, **41**:D1021-D1026.

72. Yeger-Lotem E, Riva L, Su LJ, Gitler AD, Cashikar AG, King OD, Auluck PK, Geddie ML, Valastyan JS, Karger DR, Lindquist S, Fraenkel E: **Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity**. *Nat Genet* 2009, **41**:316-323 http://dx.doi.org/10.1038/ng.337.

73. Kim YA, Przytycki JH, Wuchty S, Przytycka TM: **Modeling information flow in biological networks**. *Phys Biol* 2011, **8**:035012 http://dx.doi.org/10.1088/1478-3975/8/3/035012.

74. Huang SC, Fraenkel E: **Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks**. *Sci Signal* 2009, **2**:ra40.

75. Yeang CH, Ideker T, Jaakkola T: **Physical network models**. *J Comput Biol* 2004, **11**:243-262.

76. Gitter A, Bar-Joseph Z: **Identifying proteins controlling key disease signaling pathways**. *Bioinformatics* 2013, **29**:i227-i236 http://dx.doi.org/10.1093/bioinformatics/btt241.

77. Chasman D, Gancarz B, Hao L, Ferris M, Ahlquist P, Craven M: **Inferring host gene subnetworks involved in viral replication**. *PLoS Comput Biol* 2014, **10**:e1003626 http://dx.doi.org/10.1371/journal.pcbi.1003626.

78. Chasman D, Ho YH, Berry DB, Nemec CM, MacGilvray ME, Hose J, Merrill AE, Lee MV, Will JL, Coon JJ, Ansari AZ, Craven M, Gasch AP: **Pathway connectivity and signaling coordination in the yeast stress-activated signaling network**. *Mol Syst Biol* 2014, **10**:759.

79. Jain S, Gitter A, Bar-Joseph Z: **Multitask learning of signaling and regulatory networks with application to studying human response to flu**. *PLoS Comput Biol* 2014, **10**:e1003943 http://dx.doi.org/10.1371/journal.pcbi.1003943.

80. Gitter A, Braunstein A, Pagnani A, Baldassi C, Borgs C, Chayes J, Zecchina R, Fraenkel E: **Sharing information to reconstruct patient-specific pathways in heterogeneous diseases**. *Pac Symp Biocomput* 2014:39-50.

81. Mazza A, Gat-Viks I, Farhan H, Sharan R: **A minimum-labeling approach for reconstructing protein networks across multiple conditions**. *Algorithms Mol Biol* 2014, **9**:1 http://dx.doi.org/10.1186/1748-7188-9-1.

82. Gat-Viks I, Tanay A, Raijman D, Shamir R: **A probabilistic methodology for integrating knowledge and experiments on biological networks**. *J Comput Biol* 2006, **13**:165-181 http://dx.doi.org/10.1089/cmb.2006.13.165.

83. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM: **Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using paradigm**. *Bioinformatics* 2010, **26**:i237-i245 http://dx.doi.org/10.1093/bioinformatics/btq182.

84. Wilentzik R, Gat-Viks I: **A statistical framework for revealing signaling pathways perturbed by DNA variants**. *Nucleic Acids Res* 2015, **43**:e74 http://dx.doi.org/10.1093/nar/gkv203.

85. Bertrand D, Chng KR, Sherbaf FG, Kiesel A, Chia BKH, Sia YY, Huang SK, Hoon DSB, Liu ET, Hillmer A, Nagarajan N: **Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles**. *Nucleic Acids Res* 2015, **43**:e44 http://dx.doi.org/10.1093/nar/gku1393.

86. Shrestha R, Hodzic E, Yeung J, Wang K, Sauerwald T, Dao P, Anderson S, Beltran H, Rubin MA, Collins CC *et al.*: **HIT'nDRIVE: multi-driver gene prioritization based on hitting time**. *Research in Computational Molecular Biology*. Springer; 2014:: 293-306.

87. Vandin F, Upfal E, Raphael BJ: **Algorithms for detecting significantly mutated pathways in cancer**. *Research in Computational Molecular Biology*. Springer; 2010:: 506-521.

88. Kim YA, Cho DY, Dao P, Przytycka TM: **MEMCover: integrated analysis of mutual exclusivity and functional network reveals dysregulated pathways across multiple cancer types**. *Bioinformatics* 2015, **31**:i284-i292 http://dx.doi.org/10.1093/bioinformatics/btv247.

89. Hormozdiari F, Penn O, Borenstein E, Eichler EE: **The discovery of integrated gene networks for autism and related disorders**. *Genome Res* 2015, **25**:142-154 http://dx.doi.org/10.1101/gr.178855.114.

90. Peng CH, Jiang YZ, Tai AS, Liu CB, Peng SC, Liao CT, Yen TC, Hsieh WP: **Causal inference of gene regulation with subnetwork assembly from genetical genomics data**. *Nucleic Acids Res* 2014, **42**:2803-2819 http://dx.doi.org/10.1093/nar/gkt1277.

91. Bickmore WA: **The spatial organization of the human genome**. *Annu Rev Genomics Hum Genet* 2013, **14**:67-84.

92. Califano A, Butte AJ, Friend S, Ideker T, Schadt E: **Leveraging models of cell regulation and GWAS data in integrative network-based association studies**. *Nat Genet* 2012, **44**:841-847 http://dx.doi.org/10.1038/ng.2355.

93. Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S: **Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases**. *Nat Methods* 2016, **13**:366-370 http://dx.doi.org/10.1038/nmeth.3799.

94. Moreau Y, Tranchevent LC: **Computational tools for prioritizing candidate genes: boosting disease gene discovery**. *Nat Rev Genet* 2012, **13**:523-536 http://dx.doi.org/10.1038/nrg3253.

95. Barabási AL, Oltvai ZN: **Network biology: understanding the cell's functional organization**. *Nat Rev Genet* 2004, **5**:101-113 http://dx.doi.org/10.1038/nrg1272.

96. Köhler S, Bauer S, Horn D, Robinson PN: **Walking the interactome for prioritization of candidate disease genes**. *Am J Hum Genet* 2008, **82**:949-958 http://dx.doi.org/10.1016/j.ajhg.2008.02.013.

97. Rackham OJ, Firas J, Fang H, Oates ME, Holmes ML, Knaupp AS, Suzuki H, Nefzger CM, Daub CO, Shin JW *et al.*: **A predictive computational framework for direct reprogramming between human cell types**. *Nat Genet* 2016.

98. •• Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ: **CellNet: network biology applied to stem cell engineering**. *Cell* 2014, **158**:903-915 http://dx.doi.org/10.1016/j.cell.2014.07.020.
Network biology approach for inferring cell type and tissue-specific networks, as well as prioritizing regulators for directed reprogramming of stem cells. Validated by experiments that demonstrate its use for improving reprogramming efficiency.

99. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ: **Single-cell chromatin accessibility reveals principles of regulatory variation**. *Nature* 2015, **523**:486-490.

100. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M, Weissman JS: **Genome-scale CRISPR-mediated control of gene repression and activation**. *Cell* 2014, **159**:647-661 http://dx.doi.org/10.1016/j.cell.2014.09.029.

101. Wang T, Wei JJ, Sabatini DM, Lander ES: **Genetic screens in human cells using the CRISPR-Cas9 system**. *Science* 2014, **343**:80-84 http://dx.doi.org/10.1126/science.1246981.