Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

Network types and their application in natural variation studies in plants

José M Jiménez-Gómez[1,2,*]

We are in the age of data-driven biology. Not even a decade after the invention of high-throughput sequencing technologies, there are methods that accurately monitor DNA polymorphisms, transcription profiles, methylation states, transcription factor binding sites, chromatin compactness, nucleosome positions, dynamic histone marks, and so on. We are starting to generate comparable amounts of protein or metabolite data. A key issue is how are we going to make sense of all this information. Network analysis is the most promising method to integrate, query and display large amounts of data for human interpretation. This review shortly summarizes the basic types of networks, their properties and limitations. In addition, I introduce the application of networks to the study of the molecular mechanisms behind natural phenotypic variation.

Addresses
[1] INRA – Institut National de la Recherche Agronomique, UMR 1318, Institut Jean-Pierre Bourgin, Versailles, France
[2] Max Planck Institute for Plant Breeding Research, Department of Plant Breeding and Genetics, Carl-von-Linné-Weg 10, 50829 Cologne, Germany

Corresponding author: Jiménez-Gómez, José M. (jose.jimenez-gomez@versailles.inra.fr)
* Tel.: +33 0 1 30 83 30 00.

Current Opinion in Plant Biology 2014, **18**:80–86

This review comes from a themed issue on **Genome studies and molecular genetics**

Edited by **Kirsten Bomblies** and **Olivier Loudet**

For a complete overview see the Issue and the Editorial

Available online 15th March 2014

http://dx.doi.org/10.1016/j.pbi.2014.02.010

Introduction

Variation among individuals in a specific trait, such as plant height, flowering time or susceptibility to pathogens, is often explained by a combination of mutations, molecular interactions and environmental effects. The extended use of genomic techniques in individual labs and their availability in public repositories provide outstanding opportunities to localize these mutations and understand their interactions. At the same time, it is becoming clear that dozens of candidate variants from genome-wide association studies or hundreds of differentially expressed transcripts only represent single dimensions of a much more complex figure. Therefore, our best chance to gain insight into the molecular mechanisms controlling natural phenotypic diversity is to take a step forward from the reductionist view that has been so successful in science's history and consider their composite nature. This is challenging due to the heterogeneity of genomic datasets, and the non-linear relationship between molecular events and phenotypes. Luckily, the development of methods, sometimes derived from social sciences, engineering or physics, is allowing biologists to build more accurate pictures of molecular process by compiling information from different sources and connecting the dots (Figure 1).

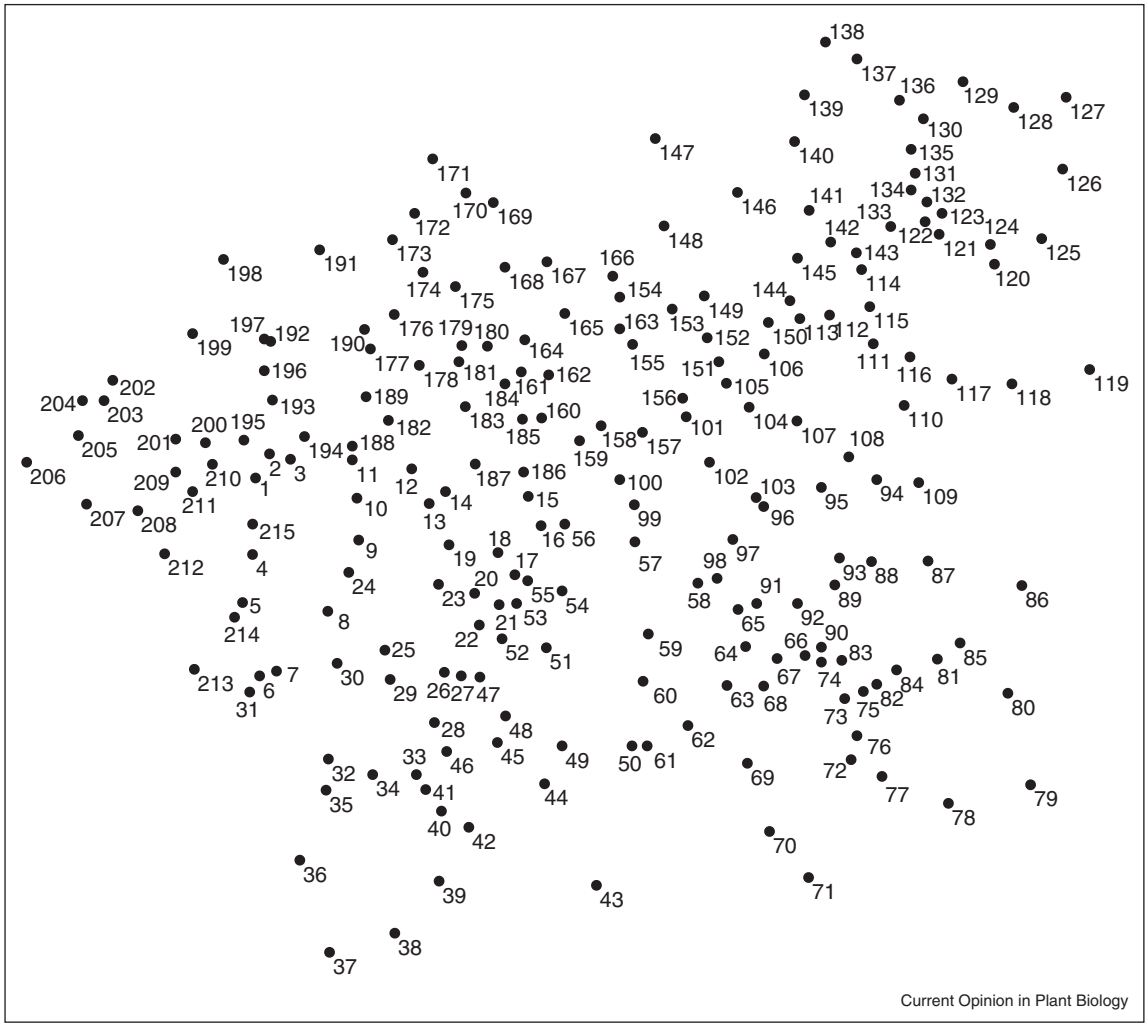
The most common representation used to describe relationships is a graph, also called network. The most abundant networks in biology have genes, proteins or metabolites represented as nodes; and genetic relationships, physical interactions or biochemical reactions as edges connecting the nodes. The study of individual nodes or node clusters in the context of these types of networks has been extremely useful for gene annotation, pathway discovery and hypothesis generation. Interpreting network topologies and node characteristics is having a great impact in our understanding of the hierarchical organization and the evolution of biological systems (Box 1). In addition to the basic types of networks, graph representations in plants have been used to represent relationships between microRNA and their targets [1,2], to describe evolutionary distances [3,4], or to filter out complex SNPs from next generation sequencing reads [5•]. Other types of data that benefit from graph representations include interactions between proteins and DNA [6] or between distant chromatin [7].

The most popular tool to represent large networks in personal computers is an open source initiative called Cytoscape (www.cytoscape.org). Cytoscape is enhanced with more than a hundred community-created plugins to easily perform topological analyses, functional enrichment, identification of sets of highly inter-connected nodes, identification of sets of nodes that show coordinated responses to stimulus, or integration of heterogeneous datasets, among others [8••]. Another open source software with infinite possibilities for network construction and analysis is R (http://www.R-project.org), although it requires users to become familiar with its programming language.

Basic types of networks

Genetic pathways are among the earliest network representations in molecular biology. These represent regulatory

Figure 1



Connecting the dots. Scientific techniques allow acquisition of molecular phenotypes at very high-throughput. Combination of this information is essential to obtain a mechanistic understanding of biological processes. For example, expression of a transcript can only be understood by integrating information from regulatory motifs, chromatin states, methylation, occupancy of transcription factors, polymerase speed, etc. In this way, generating knowledge in modern biology requires combining evidences to obtain a recognizable picture, as in a connect-the-dots game. Network analysis helps ordering the dots in the correct way.

interactions between genes or proteins, and were built comparing the phenotype of single mutants to the phenotype of double mutants [9]. Genetic networks are arguably the most helpful kind of network to understand complex links between phenotypes and genotypes. The attractiveness of these networks is that they do not attempt to represent a snapshot of an organism, but just to link functionally related molecules. Although there are no high throughput methods to build genetic networks in plants as there are in yeast [10], we can find a substitute in the multiple co-expression networks available for various plant species (e.g. [11]). These networks are frequently used for annotation of unknown genes based on the premise that genes with similar expression patterns across a number of coherent experiments are functionally related [12–14]. Recently, this 'guilt by association' principle is questioned because functional annotation seems to be encoded in very few critical interactions and it cannot be freely transmitted to the rest of the network [15•]. Another interesting use for co-expression networks when genome sequences are available is the discovery of regulatory elements in the promoters of tightly co-regulated genes [16,17].

Co-expression networks differ from genetic networks in their lack of directionality, making impossible to establish the hierarchy of its members. Dozens of algorithms, mostly based on Bayesian Networks and Gaussian Graphical Models, have been developed to predict

Box 1 Networks properties

The general topology of biological networks, as well as single node characteristics, can be described through graph theory measurements [66•]. Biological networks are organized in modules and follow a scale-free distribution, in which a reduced number of nodes, called hubs, contain most of the interactions [67]. These properties could have arisen from genome duplications, in which well-connected molecules are more likely to double their number of links than unconnected molecules. Consistent with this, older proteins typically show more connections than their younger counterparts [68]. Another characteristic of biological networks is the 'small world' property, in which most nodes in the graph are unconnected to each other, but there is a relatively short path connecting any pair of nodes. This property is shown to enhance signal propagation and synchronizability [69]. A characteristic of interest for nodes is their degree, or number of connections. It is thought that nodes with many connections tend to be essential for the network, but this is shown to not always be true [70]. For example, essentiality in metabolic networks comes from enzymes that are specific for a single reaction, while generalist enzymes that perform multiple reactions are dispensable [71,72•]. A good metric for essentiality is betweenness centrality, or the number of shortest paths passing through a node. Nodes with high betweenness centrality are more influential for the rest of the network independently of their number of connections [73]. The position of a node in a network is also a good indicative of its essentiality [74].

directional genetic interactions from transcriptome datasets (reviewed in [18]). The 5th iteration of the DREAM project evaluated 35 different methods to infer cause–effect networks from expression data and showed that a consensus based in multiple algorithms is the most accurate approach [19•]. Discerning regulators from regulated molecules at global scale is a great step towards understanding the pathways and main players controlling phenotypic variability in nature. In the future, we expect more complex models that can have into account multiple transcription factors controlling the expression of one gene, polymorphisms in regulatory motifs, or smallRNA regulation. As an example, researchers started including a time lag between transcription factors and their targets to search for novel target molecules [20].

Another common type of graphs are protein–protein interaction networks, although their utility to associate changes in complex phenotypes to molecules is questionable. The main problem with constructing protein–protein interaction (PPI) networks in plants is the low specificity and sensitivity of the available methods [21•]. Technical advances have only partially reduced these problems, indicated by the marginal overlap in the results obtained with different techniques [22,23]. These networks can also be constructed through predicted interactions that are based in functional annotation, homology to interactions in other organisms, co-localization, co-expression or literature mining (e.g. [24]). Similarly to experiment-based networks, prediction-based networks also show biases, so the former are usually preferred [21•]. The largest PPI network in plants queried 30% of known genes, leaving behind a datasets of irrefutable interest for biology [25••]. The impossibility to repeat this for each tissue, species, accession and mutant, leaves us anticipating the unfolding of high throughput sequencing for proteins [26].

A different type of graph uses metabolites as nodes. As with co-expression networks, metabolic networks can connect nodes based on co-occurrence across experiments. Recently, these networks have been useful in the analysis of natural variation in metabolite content in tomato and Arabidopsis (e.g. [27,28]). Another class of metabolic network, called metabolic models, has edges representing the biochemical reaction that transforms one metabolite into another. These are powerful tools that can predict the effect of genotype or environment in metabolite concentration by fitting mathematical algorithms that simulate the flow of metabolites through the network [29]. Metabolite network models can be build using only genome sequence and annotation [30], but removing artifacts and validating predictions require several years [31]. A number of flux-balance models are available for plant species (reviewed in [32]). Biochemical reactions in these models are still far from complete because of their complexity, their subcellular compartmentalization and their tissue and condition dependency [32,33]. Current efforts try to update metabolic models to include these factors [34].

Biases, filtering and integration

The interpretation of biological processes through analysis of networks is not straightforward. A big problem is to have into account the biases introduced by the techniques used to generate the data and the specific characteristics of each molecule. For example, if we rank the importance of a protein based on the number of neighbors it has in a PPI network, we need to consider that soluble proteins show more protein–protein interactions than membrane proteins [35]. Another problem is that abundant transcripts; proteins or metabolites are more visible to all techniques, therefore more likely to be found in networks. As mentioned above, different protocols will have preferences for specific molecules, and in fact, networks constructed from datasets obtained with different methods often have different topologies [36]. Over-representation of known interactions is a significant problem with techniques based on previous knowledge, such as microarrays, chromatin immunoprecipitation, and yeast-two-hybrid. One more limitation of network representations is their inability to represent dynamic processes. This is important in light of the current evidences showing that molecular interactions are extensively rewired depending on growth conditions, cellular compartments, cell types or environmental stresses [37]. Acknowledging these biases is important to

design the null models to derive conclusions from network analyses.

The most dangerous problem for researchers in network analyses is their low specificity, in other words, the large number of false connections they contain [38••]. Edges in biological networks have a probabilistic nature that is often not considered, and threshold variation affects network topology dramatically. There are two main strategies to increase edge specificity: filtering noise by construction of networks that are specific to the studied process and enhancing true relationships by accumulation of evidences from heterogeneous datasets.

The need for filtering false positives is shifting the initial attempts to build networks containing as many molecules as possible from an organism towards construction of more specialized networks. A general method to filter noise uses pre-clustering algorithms to define coherent expression datasets prior to network construction [39]. An excellent example of the power behind filtering is the reconstruction of the metabolic pathway producing steroidal glycoalkaloids starting from a co-expression analysis only with data from two of the species that contain this toxic compound [40•]. The popularization of genomic technologies makes possible the generation of networks from tissue and condition specific datasets. A recent work re-constructs part of the signaling network downstream of a receptor protein by using genome wide expression data from a mutant on this receptor grown under conditions known to stimulate the signals [41]. Finally, another way to focus on reconstruction of particular pathways is to link genes by their coordinated responses to a stimulus, independently of whether these genes co-express or not [42,43].

Integrating evidences from multiple sources can also help reducing the number of false positive interactions in a network. This can be challenging when combining unrelated datasets describing very different aspects of a molecule. Bayesian methods can solve this problem and have been used in Arabidopsis to enhance a protein–protein interaction network with evidences such as co-expression, annotation, sequence composition or phylogeny [44,45]. A different approach to heterogeneous data integration is to assign a common scoring scheme to all datasets and use these scores to calculate the likelihood of an interaction. This has been successful to link genes with phenotypic traits using an Arabidopsis network that integrates co-expression, protein–protein interactions in multiple organisms, protein homology, or co-citation in the literature, among others [46•].

Network analyses to elucidate the molecular mechanism underlying phenotypic variation

Network analyses allow detection of orchestrated changes that are not visible by looking at the individual molecules. This is helpful to understand global molecular changes leading natural phenotypic diversity. For example, comparison of co-expression networks between cultivated tomato and a wild relative adapted to a desertic environment, showed extensive rewiring of light regulated and photosynthesis related genes [47]. Lack of conservation in co-expression patterns has also been observed between rice and Arabidopsis in stress responsive genes, while connections between tissue-specific genes were much more conserved [48]. Analysis of multiple network types showed that rewiring occurs fastest in transcriptional regulation networks, and then in decreasing order, in genetic interactions, protein interactions, and metabolic networks [49]. For the interested reader, excellent reviews have dealt with various aspects of network evolution in plants [50,51].

It is also possible to focus in individual pathways to find associations to natural phenotypic variation. For example, variation of expression in predefined sets of genes was studied in a segregating population of Arabidopsis [52]. Also in Arabidopsis, a regulatory network controlling flowering time was inferred from a list of genes known to affect this trait, their expression patterns in a segregating population and eQTL data [53]. An interesting work in yeast proposes that lineage-specific selection in a pathway can be detected by counting the number of eQTLs with effects in the same direction among its members [54]. A very popular protocol to find pathways underlying phenotypic differences is to define sets of interconnected nodes that coordinately change across genotypes or conditions (called active modules). This method is implemented in Cytoscape and is widely used in humans to infer the regulatory networks underlying complex diseases [8••]. Changes in expression of groups of genes can also serve as markers to classify phenotypes or genotypes. These groups of genes are called set covers, and helped comparing developmental stages in meristems from distant tomato species [55].

Another popular use of networks is to prioritize candidate genes from the large lists generated in natural variation studies. In plants, co-expression networks have been used to propose candidates for associations in GWA studies in Arabidopsis [56] or for eQTL hotspots in Populus [57]. Combination of evidences from genomic variants, eQTLs, annotation and co-expression pointed the gene underlying a QTL in Arabidopsis [58]. There are a number of exciting examples outside plants where the genes underlying eQTL hotspots are identified by integrating multiple heterogeneous datasets such as co-expression, annotation, presence of regulatory motifs, protein–protein interactions, among others (e.g. [59,60••]). Moreover, recent works targeting genes underlying complex human diseases show the benefit of using genome-wide maps of variation in regulatory elements to prioritize candidate mutations [61,62••]. Genome wide maps of regulatory elements

can be now obtained with techniques that do not require previous knowledge of the motifs involved, such as DNase-seq or MNase-seq [63,64]. Using networks to integrate these types of datasets and boost the noise-to-signal ratio is perhaps the key to associate mutations in regulatory regions with phenotypic differences among related individuals.

Conclusions

Technical advances in genomics create multiple layers of valuable information. Integrating and displaying this heterogeneous information in networks will help us understand the molecular organization of organisms and their evolution. Moreover, network analyses enable generation of precise, data-driven hypothesis about the molecular changes that lead to phenotypic variation. While integrating information in networks is not the solution to all questions in biology, it may be our best bet to deal efficiently with large amounts of data and to make it work towards solving some of the standing questions and challenges in modern biology [65••].

Acknowledgements

Thanks to Eva Maria Willing for her ideas on how to illustrate this article.

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. Ma X, Shao C, Wang H, Jin Y, Meng Y: **Construction of small RNA-mediated gene regulatory networks in the roots of rice (*Oryza sativa*)**. *BMC Genomics* 2013, **14**:510.

2. Meng Y, Shao C, Gou L, Jin Y, Chen M: **Construction of MicroRNA- and MicroRNA*-mediated regulatory networks in plants**. *RNA Biol* 2011, **8**:1124-1148.

3. Alvarez-Ponce D, Lopez P, Bapteste E, McInerney JO: **Gene similarity networks provide tools for understanding eukaryote origins and evolution**. *Proc Natl Acad Sci* 2013, **110**:E1594-E1603.

4. Li S, Pandey S, Gookin TE, Zhao Z, Wilson L, Assmann SM: **Gene-sharing networks reveal organizing principles of transcriptomes in Arabidopsis and other multicellular organisms**. *Plant Cell* 2012, **24**:1362-1378.

5. • Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE: **Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol**. *PLoS Genet* 2013, **9**:e1003215.
This is a network approach to identify a reliable set of SNPs in heterozygous polyploid species without a reference genome. Nodes represent reads and edges link polymorphic reads. The method filters out complex variants based on the topology of the sub-networks.

6. Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, Sweet L, Gomes A, Rustad T, Dolganov G, Glotova I *et al.*: **The *Mycobacterium tuberculosis* regulatory network and hypoxia**. *Nature* 2013, **499**:178-183.

7. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO *et al.*: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome**. *Science* 2009, **326**:289-293.

8. •• Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T: **A travel guide to Cytoscape plugins**. *Nat Methods* 2012, **9**:1069-1076.
Cytoscape is the most popular software for network construction, display and analysis; but the number of plugins available is overwhelming. This guide does a great job directing you to the best plugins based on their utility and popularity.

9. Koornneef M, Hanhart CJ, van der Veen JH: **A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana***. *Mol Gen Genet* 1991, **229**:57-66.

10. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S *et al.*: **The genetic landscape of a cell**. *Science* 2010, **327**:425-431.

11. Yim WC, Yu Y, Song K, Jang CS, Lee B-M: **PLANEX: the plant co-expression database**. *BMC Plant Biol* 2013, **13**:83.

12. Bassel GW, Glaab E, Marquez J, Holdsworth MJ, Bacardit J: **Functional network construction in Arabidopsis using rule-based machine learning on large-scale data sets**. *Plant Cell* 2011, **23**:3101-3116.

13. Childs KL, Davidson RM, Buell CR: **Gene coexpression network analysis as a source of functional annotation for rice genes**. *PLoS ONE* 2011, **6**:e22196.

14. Fukushima A, Nishizawa T, Hayakumo M, Hikosaka S, Saito K, Goto E, Kusano M: **Exploring tomato gene functions based on coexpression modules using graph clustering and differential coexpression approaches**. *Plant Physiol* 2012, **158**:1487-1502.

15. • Gillis J, Pavlidis P: **'Guilt by association' is the exception rather than the rule in gene networks**. *PLoS Comput Biol* 2012, **8**:e1002444.
Gene annotation using networks is very useful for non-model organisms. This paper shows that one must be careful in this process, as freely spreading functional annotation through the network will frequently result in false positive annotations.

16. Ruan J, Perez J, Hernandez B, Lei C, Sunter G, Sponsel VM: **Systematic identification of functional modules and cis-regulatory elements in *Arabidopsis thaliana***. *BMC Bioinform* 2011, **12(Suppl 12)**:S2.

17. Zheng X, Liu T, Yang Z, Wang J: **Large cliques in Arabidopsis gene coexpression network and motif discovery**. *J Plant Physiol* 2011, **168**:611-618.

18. De Smet R, Marchal K: **Advantages and limitations of current network inference methods**. *Nat Rev Microbiol* 2010, **8**:717-729.

19. • Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, The DREAM5 Consortium, Kellis M, Collins JJ *et al.*: **Wisdom of crowds for robust gene network inference**. *Nat Methods* 2012, **9**:796-804.
Compilation of the results of a contest between methods to construct gene regulatory networks. Different algorithms complement each other and their limitations are canceled out, so integrating their results is the most accurate.

20. Srivastava GP, Li P, Liu J, Xu D: **Identification of transcription factor's targets using tissue-specific transcriptomic data in *Arabidopsis thaliana***. *BMC Syst Biol* 2010, **4(Suppl 2)**:S2.

21. • Braun P, Aubourg S, Van Leene J, De Jaeger G, Lurin C: **Plant protein interactomes**. *Annu Rev Plant Biol* 2013, **64**:161-187.
Fantastic review about possibilities and limitations of current methods to detect and analyze protein–protein interactions in plants.

22. Boruc J, Van den Daele H, Hollunder J, Rombauts S, Mylle E, Hilson P, Inzé D, De Veylder L, Russinova E: **Functional modules in the Arabidopsis core cell cycle binary protein–protein interaction network**. *Plant Cell* 2010, **22**:1264-1280.

23. Van Leene J, Boruc J, De Jaeger G, Russinova E, De Veylder L: **A kaleidoscopic view of the Arabidopsis core cell cycle interactome**. *Trends Plant Sci* 2011, **16**:141-150.

24. Lin M, Shen X, Chen X: **PAIR: the predicted Arabidopsis interactome resource**. *Nucleic Acids Res* 2011, **39**:D1134-D1140.

25. •• Consortium AIM: **Evidence for network evolution in an Arabidopsis interactome map**. *Science* 2011, **333**:601-607.
This is the largest protein–protein interaction network in plants. New proteins are assigned to known processes such as ubiquitination or auxin signaling. Moreover, the authors study the rewiring of interactions after genome duplication.

26. Nivala J, Marks DB, Akeson M: **Unfoldase-mediated protein translocation through an α-hemolysin nanopore**. *Nat Biotechnol* 2013, **31**:247-250.

27. Toubiana D, Semel Y, Tohge T, Beleggia R, Cattivelli L, Rosental L, Nikoloski Z, Zamir D, Fernie AR, Fait A: **Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations**. *PLoS Genet* 2012, **8**:e1002612.

28. Sulpice R, Nikoloski Z, Tschoep H, Antonio C, Kleessen S, Larhlimi A, Selbig J, Ishihara H, Gibon Y, Fernie AR *et al.*: **Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of Arabidopsis accessions**. *Plant Physiol* 2013, **162**: 347-363.

29. Simons M, Misra A, Sriram G: **Genome-scale models of plant metabolism**. *Methods Mol Biol Clifton, NJ* 2014, **1083**:213-230.

30. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: **High-throughput generation, optimization and analysis of genome-scale metabolic models**. *Nat Biotechnol* 2010, **28**:977-982.

31. Thiele I, Palsson BØ: **A protocol for generating a high-quality genome-scale metabolic reconstruction**. *Nat Protoc* 2010, **5**:93-121.

32. Sweetlove LJ, Ratcliffe RG: **Flux-balance modeling of plant metabolism**. *Front Plant Physiol* 2011, **2**:38.

33. Heinig U, Gutensohn M, Dudareva N, Aharoni A: **The challenges of cellular compartmentalization in plant metabolic engineering**. *Curr Opin Biotechnol* 2013, **24**:239-246.

34. Mintz-Oron S, Meir S, Malitsky S, Ruppin E, Aharoni A, Shlomi T: **Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity**. *Proc Natl Acad Sci U S A* 2012, **109**:339-344.

35. Yu H, Zhu X, Greenbaum D, Karro J, Gerstein M: **TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics**. *Nucleic Acids Res* 2004, **32**:328-337.

36. Giorgi FM, Del Fabbro C, Licausi F: **Comparative study of RNA-seq- and microarray-derived coexpression networks in *Arabidopsis thaliana***. *Bioinforma Oxf Engl* 2013, **29**:717-724.

37. Bandyopadhyay S, Mehta M, Kuo D, Sung MK, Chuang R, Jaehnig EJ, Bodenmiller B, Licon K, Copeland W, Shales M *et al.*: **Rewiring of genetic networks in response to DNA damage**. *Science* 2010, **330**:1385-1389.

38. •• Ideker T, Dutkowski J, Hood L: **Boosting signal-to-noise in complex biology: prior knowledge is power**. *Cell* 2011, **144**:860-863.

This article gives four examples of how to enhance the extraction of relevant information from complex datasets. In all cases the key is the use of filters and integrators.

39. Feltus FA, Ficklin SP, Gibson SM, Smith MC: **Maximizing capture of gene co-expression relationships through pre-clustering of input expression samples: an Arabidopsis case study**. *BMC Syst Biol* 2013, **7**:44.

40. • Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R *et al.*: **Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes**. *Science* 2013, **341**:175-179.

Using the expression of a single gene, the authors fish out co-expressed partners to completely re-construct the metabolic pathway leading to production of a toxic compound in solanaceas.

41. Vidal EA, Moyano TC, Riveras E, Contreras-López O, Gutiérrez RA: **Systems approaches map regulatory networks downstream of the auxin receptor AFB3 in the nitrate response of *Arabidopsis thaliana* roots**. *Proc Natl Acad Sci U S A* 2013, **110**:12840-12845.

42. Less H, Angelovici R, Tzin V, Galili G: **Coordinated gene networks regulating Arabidopsis plant metabolism in response to various stresses and nutritional cues**. *Plant Cell* 2011, **23**:1264-1271.

43. Maere S, Dijck PV, Kuiper M: **Extracting expression modules from perturbational gene expression compendia**. *BMC Syst Biol* 2008, **2**:33.

44. Cui J, Li P, Li G, Xu F, Zhao C, Li Y, Yang Z, Wang G, Yu Q, Li Y *et al.*: **AtPID: *Arabidopsis thaliana* protein interactome database – an integrative platform for plant systems biology**. *Nucleic Acids Res* 2008, **36**:D999-D1008.

45. Cui J, Liu J, Li Y, Shi T: **Integrative identification of Arabidopsis mitochondrial proteome and its function exploitation through protein interaction network**. *PLoS ONE* 2011, **6**:e16022.

46. • Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY: **Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana***. *Nat Biotechnol* 2010, **28**: 149-156.

The network constructed here is an impressive example of data integration. Datasets obtained with multiple techniques and encompassing all kingdoms of life are used to detect evidences of functional relationship between Arabidopsis genes.

47. Koenig D, Jimenez-Gomez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat AV *et al.*: **Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato**. *Proc Natl Acad Sci U S A* 2013, **110**:E2655-E2662.

48. Movahedi S, Peer YV, de Vandepoele K: **Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in arabidopsis and rice**. *Plant Physiol* 2011, **156**:1316-1330.

49. Shou C, Bhardwaj N, Lam HYK, Yan K-K, Kim PM, Snyder M, Gerstein MB: **Measuring the evolutionary rewiring of biological networks**. *PLoS Comput Biol* 2011, **7**:e1001050.

50. Chae L, Lee I, Shin J, Rhee SY: **Towards understanding how molecular networks evolve in plants**. *Curr Opin Plant Biol* 2012, **15**:177-184.

51. De Smet R, Van de Peer Y: **Redundancy and rewiring of genetic networks following genome-wide duplication events**. *Curr Opin Plant Biol* 2012, **15**:168-176.

52. Kliebenstein DJ, West MA, Leeuwen H, van Loudet O, Doerge RW, Clair DAS: **Identification of QTLs controlling gene expression networks defined a priori**. *BMC Bioinform* 2006, **7**:1-17.

53. Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC: **Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci**. *Proc Natl Acad Sci U S A* 2007, **104**: 1708-1713.

54. Fraser HB, Moses AM, Schadt EE: **Evidence for widespread adaptive evolution of gene expression in budding yeast**. *Proc Natl Acad Sci* 2010 http://dx.doi.org/10.1073/pnas.0912245107.

55. Park SJ, Jiang K, Schatz MC, Lippman ZB: **Rate of meristem maturation determines inflorescence architecture in tomato**. *Proc Natl Acad Sci U S A* 2012, **109**:639-644.

56. Chan EKF, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ: **Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana***. *PLoS Biol* 2011, **9**:e1001125.

57. Drost DR, Benedict CI, Berg A, Novaes E, Novaes CRDB, Yu Q, Dervinis C, Maia JM, Yap J, Miles B *et al.*: **Diversification in the genetic architecture of gene expression and transcriptional networks in organ differentiation of Populus**. *Proc Natl Acad Sci* 2010, **107**:8492-8497.

58. Jiménez-Gómez JM, Wallace AD, Maloof JN: **Network analysis identifies ELF3 as a QTL for the shade avoidance response in Arabidopsis**. *PLoS Genet* 2010, **6**:e1001100.

59. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, Bumgarner RE, Schadt EE: **Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks**. *Nat Genet* 2008, **40**:854-861.

60. •• Zhu J, Sova P, Xu Q, Dombek KM, Xu EY, Vu H, Tu Z, Brem RB, Bumgarner RE, Schadt EE: **Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation**. *PLoS Biol* 2012, **10**:e1001301.
This is a nice example of how integration of datasets can greatly help the identification of genes underlying phenotypic variation, in this case gene expression and metabolite abundance.

61. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al.*: **Mapping and analysis of chromatin state dynamics in nine human cell types**. *Nature* 2011, **473**:43-49.

62. •• Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J *et al.*: **Systematic localization of common disease-associated variation in regulatory DNA**. *Science* 2012, **337**:1190-1195.
This work shows that genome-wide information about regulatory regions is a very powerful predictor when integrated with GWAS results in humans.

63. Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S: **Epigenome characterization at single base-pair resolution**. *Proc Natl Acad Sci* 2011 http://dx.doi.org/10.1073/pnas.1110731108.

64. Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS *et al.*: **Global mapping of protein–DNA interactions in vivo by digital genomic footprinting**. *Nat Methods* 2009, **6**:283-289.

65. •• Pacific Biosciences. The New Biology. http://www.youtube.com/watch?v=sjTQD6E3lH4&feature=youtu.be. 2011. This documentary gathers interviews with the leading scientists in the field of genomics and fantastic examples of how network analyses can help modern society. A must watch.

66. • Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, Schneider R, Bagos PG: **Using graph theory to analyze biological networks**. *BioData Min* 2011, **4**:10.
This is a very complete review on what parameters can be measured in biological networks and what they mean for a biologist.

67. Barabási A-L, Albert R: **Emergence of scaling in random networks**. *Science* 1999, **286**:509-512.

68. Fokkens L, Hogeweg P, Snel B: **Gene duplications contribute to the overrepresentation of interactions between proteins of a similar age**. *BMC Evol Biol* 2012, **12**:99.

69. Watts DJ, Strogatz SH: **Collective dynamics of ''small-world'' networks**. *Nature* 1998, **393**:440-442.

70. Zotenko E, Mestre J, O'Leary DP, Przytycka TM: **Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality**. *PLoS Comput Biol* 2008, **4**:e1000140.

71. Hanada K, Sawada Y, Kuromori T, Klausnitzer R, Saito K, Toyoda T, Shinozaki K, Li W-H, Hirai MY: **Functional compensation of primary and secondary metabolites by duplicate genes in *Arabidopsis thaliana***. *Mol Biol Evol* 2011, **28**:377-382.

72. • Nam H, Lewis NE, Lerman JA, Lee D-H, Chang RL, Kim D, Palsson BO: **Network context and selection in the evolution to enzyme specificity**. *Science* 2012, **337**:1101-1104.
Most reactions in a cell are performed by a small number of enzymes that are not essential for the organism and have low specificity. In contrast, there is a set of specialist, essential enzymes that catalyze single reactions.

73. Roy S: **Systems biology beyond degree, hubs and scale-free networks: the case for multiple metrics in complex networks**. *Syst Synth Biol* 2012, **6**:31-34.

74. Khurana E, Fu Y, Chen J, Gerstein M: **Interpretation of genomic variants using a unified biological network approach**. *PLoS Comput Biol* 2013, **9**:e1002886.