ELSEVIER

Development through the eyes of functional genomics

Andrew G Fraser[1] and Edward M Marcotte[2]

In many of the model organisms used to study development, it is becoming relatively routine to carry out global analyses of gene function. These analyses take many forms, from microarray analyses to the construction of physical interaction maps to the systematic analyses of loss-of-function phenotypes. Such large-scale datasets can be integrated to generate complex gene networks, and we explore how these gene networks can contribute to an understanding of developmental pathways. In particular, we examine how combining large-scale expression experiments and gene networks may move us towards a molecular description of the events of development, embodied in a succession of stage-specific subnetworks sampled from an organism's overall gene network.

Addresses
[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
e-mail: agf@sanger.ac.uk
[2]Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas, 78712, USA
e-mail: marcotte@icmb.utexas.edu

Current Opinion in Genetics & Development 2004, **14**:336–342

This review comes from a themed issue on
Pattern formation and developmental mechanisms
Edited by Derek Stemple and Jean-Paul Vincent

Available online 24th June 2004

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2004.06.015

Abbreviations
ChIP chromatin immunoprecipitation
RNAi RNA interference

Introduction

The molecular analysis of development has traditionally focussed particularly on the identification and analysis of 'master regulators' — the transcription factors and signalling pathways that control development. For example, mis-expression of the *Drosophila* homeobox gene *Antennapedia* can convert antennae to legs, a complete switch of developmental programme [1]. The identification of such key regulators, and their ordering into complex hierarchies (e.g. see [2–6]) tells us much about why development proceeds down well-described paths and this is essential for our understanding of any metazoan. However, such control theory analyses do not tell us what genes these master regulators control, nor do they attempt to describe the detailed bulk of molecular events happening during subsequent development.

Functional genomics techniques like microarray analysis of gene expression and systematic mapping of physical interactions allow us to survey more comprehensively than ever before the precise events that occur during many biological processes including development. In this review, we examine how a careful and comprehensive description of the molecular networks arising during development can complement the control-based view.

Integrating diverse datasets creates probabilistic gene networks

Functional genomics tools offer powerful ways to explore biological processes. Each dataset generated is the result of a directed experiment and suggests possible testable hypotheses: for example, one might ask which genes are induced following DNA damage [7,8]; those genes might be involved in DNA repair and checkpoint control. When many such large-scale datasets exist, rather than querying them individually, one can begin to identify genes that behave similarly across multiple datasets. For example, one can cluster genes together on the basis of similarities in expression profiles across multiple microarray experiments [9,10]. The underlying assumption of integrative biology is that genes that cluster together in this way share biological function [11]. One can thus use this approach either to assign function to previously uncharacterised genes on the basis of their clustering with well-known genes (the so-called 'guilt by association' method; e.g. see [12]), or to discover new processes through the identification of highly-clustered genes of unknown molecular function. One can even search for genes that co-express and whose orthologs in another organism also co-express — genes that meet this stringent criterion are presumably even more likely to share biological function [13•].

Just as there are many ways of examining gene function on a large scale, so there are many ways to use large-scale datasets to define links between genes that may reflect shared function. For example, genes can be linked by virtue of large-scale assays measuring physical interactions between their protein products [14–20,21••,22••,23,24]. They may be linked by similarities in their mRNA expression profiles [9,13•,25,26•,27–29]. They may also be linked by genetic interactions [30,31•], by shared subcellular localisation [32•,33,34•], by similarities in their loss-of-function phenotypes [35,36•,37–39], and by a host of bioinformatics analyses (reviewed in [40–43]). Each set of linkages is based on a particular facet of gene function,

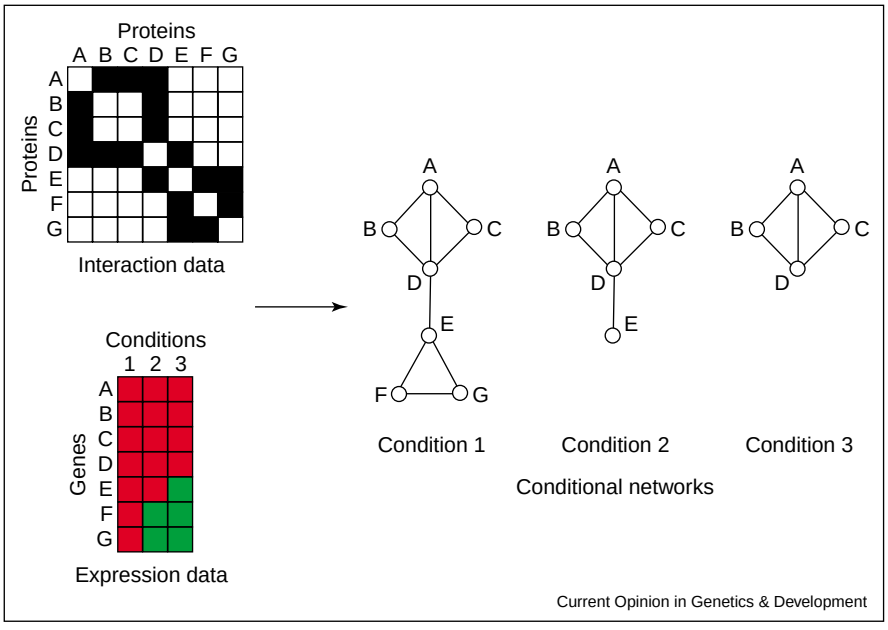
and thus each yields complementary sets of links between genes. Various statistical methods can be used to integrate these diverse datasets such that the more datasets one integrates, the more functionally significant the gene linkages become [34•,44]. In these statistical frameworks, noise cancels and signal adds, resulting in networks that improve in confidence and coverage as new datasets are added. This approach yields very large and complex networks of gene linkages — for example, even networks based only on physical interactions of *Saccharomyces cerevisiae* proteins comprise >50% of all yeast genes with an average of 2–3 linkages per gene [45•].

A complex gene network, in these terms, is not a simple physical entity. Instead, it is the complete set of statistically significant links between genes that can be identified from experimental data. Any individual link between two genes may comprise genetic interaction data, expression correlations, physical interactions and so on; taken together, there is significant evidence that the two genes are somehow functionally related. (We discuss such 'probabilistic gene networks' at length in [45•].) Examining groups of genes that are linked functionally allows the researcher to identify the systems and machineries that run the cell [29,46] — the ribosome, the spliceosome, and so on — as well as identify novel groupings of genes of unknown (but related) function [47]. In all cases, the underlying assumption in using these networks to describe and understand gene function is that closely linked genes are likely to share functions.

Development as a coordinated series of subnetworks

It is clear that what occurs in a particular cell at any one time is different to the full gene linkages shown in such global networks. Take a protein–protein interaction map as perhaps the simplest example: contacts *in vivo* can obviously only be made between proteins that are present in a cell at the same time and in the same compartment. However, no such constraints are made when constructing a physical interaction map using yeast two-hybrid data. The connections seen in a complete physical interaction map for yeast or fly or worm thus form a *master network*, the total set of possible pairwise interactions between proteins encoded in those genomes; the physical interaction networks of a muscle cell, a neuron, or an oocyte are *subnetworks*, subsets of this master network. This is true for all of the global networks assembled through the continued integration of large-scale datasets in model organisms; in each case, integration of the datasets results in a master network that holds all possible gene–gene connections, whereas only a subset of these is present in any cell or tissue at any time (Figure 1). The relationship between the master network and the specific

Figure 1



Protein interaction subnetworks specific to developmental stages can be derived from the master protein interaction network and microarray data. This strategy requires a reasonably accurate protein interaction network derived from integration and filtering of large-scale experiments. Starting with such a 'master interaction network', indicated by the matrix of interactions between proteins (A–G) on the top left, proteins can be excluded based upon the corresponding genes' lack of expression (bottom left) in different cell conditions (1–3), resulting in a set of protein subnetworks specific to the cell conditions. Given a series of microarray experiments from differing cell types or developmental stages, one might derive a succession of networks, capturing the evolving gene networks controlling the developmental processes.

subnetworks is exactly analogous to that between genome and transcriptome: one represents the total potential genetic information, the other the subset active in the cell of interest.

Development can thus be thought of as a highly regulated and progressive generation of subnetworks, each corresponding to a specific cell-type or cell-state from totipotent to pluripotent to terminally differentiated. Each subnetwork is not merely the genes expressed in that cell-type, but their connections and interactions as well. Viewing development in this global way leads immediately to two questions: what are all the subnetworks that arise during development, and why does a particular subnetwork arise in a particular cell at a specific time? Here, we concentrate on the former: can we map out all the subnetworks that arise during development and, if so, what can this tell us?

Cell signatures: the ultimate developmental markers

Defining cell-types during development is traditionally done in several ways: on the basis of lineage (where did this cell derive from and what cells does it give rise to?); on the basis of anatomy (where is this cell, and what cells does it contact?); and on the basis of appearance (what does it look like and what markers does it express?) This cellular description of development has been done most completely for the worm *Caenorhabditis elegans*, where, dating from the pioneering work of Sulston and colleagues [48], every cell division, migration and cell death event is known from fertilisation through to adulthood. Even in this completely described case, while we know the lineage and fate of every somatic cell in the animal, we still understand relatively little about the molecular networks underlying this development: how many genuinely distinct cell types are there in the worm? What genes are expressed in each cell type and what regulatory circuitry defines this? What are the protein–protein interactions occurring in each different cell-type? Effectively, we would love to know as completely as possible every one of the subnetworks that arises during development and in which cells these subnetworks appear. This goal, should we reach it, might be viewed as the *complete molecular description of the events of development*, a molecular counterpart to the complete lineage.

The simplest way to begin such a vast task is to define where and when genes are expressed; clearly, if a gene is not expressed in a defined cell type, then it can play no role in that subnetwork. Such a 'gene atlas' therefore gives a good first estimate of cell-type specific subnetworks, and we illustrate a simple example of two such subnetworks in Figure 2. This ambitious plan to map out the expression patterns of (almost) every predicted gene is proceeding in several organisms ranging from the simple to the complex, from *C. elegans* to *Ciona intestinalis* to humans.

Similar in spirit to the large number of preceding SAGE (serial analysis of gene expression) and EST (expressed sequence tag) collections, several projects have used DNA microarrays to map cell- and tissue-specific mRNA expression patterns. It is hard to be comprehensive, but a few cases will serve to make the point. For example, 79 human and 61 mouse tissues were profiled transcriptionally in a gene atlas project by Hogenesch and coworkers [49], serving to define the transcriptomes associated with each of the tissues under normal growth conditions. In plants, an analogous project in tree biology involved mapping of tissue-specific transcript profiles for different developmental stages of xylogenesis [50], followed up with measurements of transcript dynamics in autumn leaf senescence, observing the shift from photosynthesis, as chlorophyll is degraded, to alternative mechanisms of energy generation [51]. In other examples, the transcriptomes of hematopietic cell populations have been mapped under resting (and in some cases stimulated) conditions, showing marked expression differences among the different cell lineages [52], and a series of *C. elegans* developmental stages and mutants have been profiled transcriptionally to identify stage-specific gene expression (e.g. see [53–55]).

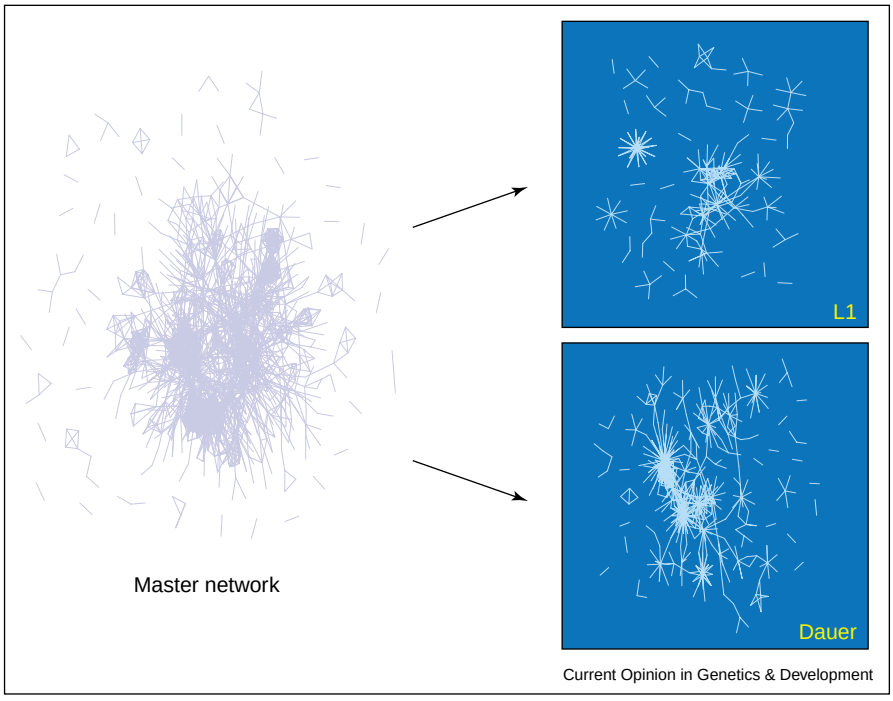
Beyond microarrays, high-throughput *in situ* hybridisation projects, often relying on imaging at the single cell level, are defining organism-wide gene expression patterns, and have made strong progress in mapping spatial expression patterns for thousands of genes in mouse [56], *Xenopus* [57], and fly embryos [58]. In principle, provided the cell types in these images can be identified unambiguously, these experiments should also provide tissue- and cell-type specific expression atlases, to be used in much the same manner we describe above.

From cell signatures to developmental mechanisms

One might be tempted to think that an attempt to define as completely as possible all the subnetworks that arise in development is little more than a cataloguing exercise. It is far more than that, however, offering not just completion but also great mechanistic insight.

Most obviously, identifying all the genes specifically expressed in a certain cell allows one to compare their putative promoters/enhancers and search for *cis*-regulatory elements that define that expression pattern. For example, Gaudet *et al.* used arrays to compare gene expression between a worm entirely lacking a pharynx and worms with excess pharangeal tissue [59]. Examination of the upstream sequences of genes that were under-expressed in the mutant lacking a pharynx identified binding sites for the transcription factor PHA-4. PHA-4 is required for development of the pharynx and the binding sites are necessary and sufficient for pharyngeal expression. Similarly, from a series of microarray

Figure 2



An example of developmental stage-specific subnetworks in *C. elegans*. On the left, we show a limited portion of a 'master' protein interaction network in *C. elegans* (derived in [45•] by 'transporting' a protein interaction network of *S. cerevisiae* genes into *C. elegans* by mapping each yeast gene in the network to its worm counterpart). In the network shown, each node is a worm gene (or in some cases several worm paralogs), and each link between genes is hypothetical, based uniquely on links between yeast genes, but hypothesised to exist in worm. The two subnetworks on the right include only those genes (and their direct interaction partners) that are upregulated in L1 worms relative to Dauer worms (labelled 'L1'), or upregulated in Dauer worms, relative to L1 worms (labelled 'Dauer') [55]. This version of the master network is quite incomplete, resulting in incomplete subnetworks, highlighting the importance of an accurate master network in this strategy.

experiments spanning the *D. melanogaster* lifecycle, genes of terminally differentiated muscle were identified, 65% of which contained binding sites for the MADS box transcription factor *dMef2* [60•]. It is these transcriptional elements that ultimately form one of the key interfaces between the regulatory circuitry of signalling pathways and transcriptional networks and the 'business end' of gene expression that defines a cell-type. Expression data like these, combined perhaps with either ChIP–chip (ChIP measured using DNA microarrays) or comparative genomics analyses is likely to greatly increase our understanding of what changes occur during differentiation of specific tissues, and also of how the complex interplay between transcriptional regulators at individual promoters brings about these effects.

Perhaps less obviously, one may be also able to use these catalogues of gene expression as sensitive tools for the interpretation of experimental data. In the same way that a specific marker can be used to define a cell-type (e.g. the use of cell surface markers to classify cells of the immune system), these subnetworks can thus serve as more comprehensive 'signatures' of individual cell-types or cell states. If we knew all the normal subnetworks that arise during development, perhaps we could rapidly analyse a perturbed network (e.g. in a mutant animal, or following a drug treatment) in terms of expansions and removals of subnetworks; one might thus infer the effect of the perturbation, in effect deconvoluting the network into its component parts, much as has been done for expression data [61].

More immediately, we can perturb development in a targeted way using a series of mutants in specific pathways and catalogue the effect of each mutation on the networks. The most immediately tractable way to do this is through array analysis of gene-expression changes. An example of this strategy is the mapping of fly-eye-specific gene expression from microarray profiles of *eyes absent* mutants [60•]. Extending this approach to a set of additional mutants would create a catalogue of perturbed networks that can serve as a reference set with which to analyse the genes of unknown function. For example, repression of the ras–raf–MAPK pathway should have a diagnostic effect on the networks that differs from perturbations in TGF-β signalling. One can then place genes of unknown function into known pathways by comparing

the effect of mutating the unknown gene with the diagnostic changes observed previously.

In *S. cerevisiae*, this approach has proved hugely successful. Hughes *et al.* profiled the expression of all yeast genes in a series of ~300 deletion mutants [62]. Clustering these data enabled the authors to identify the putative functions of several previously uncharacterised genes. More impressively, they also were able to deduce the mechanism of action of an antifungal drug through examining the effect of the drug on gene expression and comparing that to the previously compiled compendium of gene-expression profiles. One can easily imagine being able to do a similar (albeit more complex) analysis in model organisms or mammalian cell lines by examining the effect of either mutant alleles or RNAi on gene expression and assembling a similar compendium of gene-expression profiles. Such a compendium would be an excellent tool both for the potential identification of drug-action mechanisms and also for the functional characterisation of unknown genes.

It is not only expression profiles that can be gathered in this way. Synthetic Genetic Array analysis has recently proved very successful in *S. cerevisiae* at defining clusters of genes with related molecular functions [30,31•], and several groups have begun to map out systematically the synthetic lethal interactions between yeast genes. By comparing sets of synthetic lethals with synthetic phenotypes derived from deletion mutants upon drug addition, Parsons and co-workers were able to identify functions for unknown genes and suggest mechanisms of drug action [63]. Synthetic lethal analysis is being carried out in organisms ranging from worms [64] to flies to mammalian cell lines [65,66] and this type of approach is likely to be increasingly useful as the sets of data grow.

Finally, it may even be possible to extract the hierarchical organisation of 'master regulators' from systematically compiled expression data. Using sophisticated iterative clustering methods, Eran Segal and co-workers showed that it is possible to describe genome-wide expression data in yeast and mammals in terms of the expression changes and the possible logical modules of transcription factors and signalling components [67]. This approach allowed them to construct hierarchical logical architectures of these putative master regulators that are consistent with the expression data and thus to discover the regulatory circuitry underlying many processes. Analysing as comprehensively as possible the subnetworks that arise during development can thus give insight in many different ways into developmental mechanisms.

Conclusions

Describing development as a regulated progression of complex gene networks is a realisable long-term goal in several model organisms. Analysis of the subnetworks that arise and disappear at different points in development can provide an excellent complement to the more control-theory-based views of development, and may also clarify and illuminate some of the structures of the regulatory networks of master regulators that control development. We are still far from this goal at this point in any model organism but, at least in worms and flies, we already have a draft physical interaction map along with an increasingly precise set of expression data and genetic interaction data. As these datasets fill out, improving by successive approximation with each additional experiment, our view of the master network and the staged subnetworks will come into focus and provide a new and fuller framework for viewing development.

Acknowledgements

The authors acknowledge the Tx-UK Collaborative Research Initiative. AG Fraser acknowledges funding from The Wellcome Trust; EM Marcotte acknowledges funding from the National Science Foundation, National Institutes of Health, Packard Foundation, and Welch Foundation.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Schneuwly S, Klemenz R, Gehring WJ: **Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia**. *Nature* 1987, **325**:816-818.

2. Cline TW: **Autoregulation functioning of a Drosophila gene product that establish es and maintains the sexually determined state**. *Genetics* 1984, **107**:231-277.

3. Sassone-Corsi P: **CREM: a master-switch regulating the balance between differentiation and apoptosis in male germ cells**. *Mol Reprod Dev* 2000, **56**:228-229.

4. Dean Rider S Jr, Henderson JT, Jerome RE, Edenberg HJ, Romero-Severson J, Ogas J: **Coordinate repression of regulators of embryonic identity by PICKLE during germination in *Arabidopsis***. *Plant J* 2003, **35**:33-43.

5. Wang Z, Wang DZ, Pipes GC, Olson EN: **Myocardin is a master regulator of smooth muscle gene expression**. *Proc Natl Acad Sci USA* 2003, **100**:7129-7134.

6. Baudino TA, McKay C, Pendeville-Samain H, Nilsson JA, Maclean KH, White EL, Davis AC, Ihle JN, Cleveland JL: **c-Myc is essential for vasculogenesis and angiogenesis during development and tumor progression**. *Genes Dev* 2002, **16**:2530-2543.

7. Gasch AP, Huang M, Metzner S, Botstein D, Elledge SJ, Brown PO: **Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p**. *Mol Biol Cell* 2001, **12**:2987-3003.

8. Zhou Y, Gwadry FG, Reinhold WC, Miller LD, Smith LH, Scherf U, Liu ET, Kohn KW, Pommier Y, Weinstein JN: **Transcriptional regulation of mitotic genes by camptothecin-induced DNA damage: microarray analysis of dose- and time-dependent effects**. *Cancer Res* 2002, **62**:1688-1695.

9. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns**. *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.

10. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC *et al.*: **Distinctive gene expression patterns in human mammary epithelial cells and breast cancers**. *Proc Natl Acad Sci USA* 1999, **96**:9212-9217.

11. Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D: **A combined algorithm for genome-wide prediction of protein function**. *Nature* 1999, **402**:83-86.

12. Walker MG, Volkmuth W, Sprinzak E, Hodgson D, Klingler T: **Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes**. *Genome Res* 1999, **9**:1198-1203.

13. • Stuart JM, Segal E, Koller D, Kim SK: **A gene-coexpression network for global discovery of conserved genetic modules**. *Science* 2003, **302**:249-255.
The paper searches for gene pairs whose coexpression is conserved across evolution, and defines an improved worm gene network based on this conserved coexpression.

14. Fromont-Racine M, Rain JC, Legrain P: **Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens**. *Nat Genet* 1997, **16**:277-282.

15. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM *et al.*: **Functional organization of the yeast proteome by systematic analysis of protein complexes**. *Nature* 2002, **415**:141-147.

16. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K *et al.*: **Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry**. *Nature* 2002, **415**:180-183.

17. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y: **A comprehensive two-hybrid analysis to explore the yeast protein interactome**. *Proc Natl Acad Sci USA* 2001, **98**:4569-4574.

18. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y: **Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins**. *Proc Natl Acad Sci USA* 2000, **97**:1143-1147.

19. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P *et al.*: **A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae**. *Nature* 2000, **403**:623-627.

20. Tong AH, Drees B, Nardelli G, Bader GD, Brannetti B, Castagnoli L, Evangelista M, Ferracuti S, Nelson B, Paoluzi S *et al.*: **A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules**. *Science* 2002, **295**:321-324.

21. •• Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E *et al.*: **A protein interaction map of *Drosophila melanogaster***. *Science* 2003, **302**:1727-1736.
First large-scale yeast two-hybrid study of *Drosophila* protein interactions, offering an unprecedented view of metazoan gene networks.

22. •• Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T *et al.*: **A map of the interactome network of the metazoan *C. elegans***. *Science* 2004, **303**:540-543.
First large-scale yeast two-hybrid analysis of *C. elegans* proteins, and another important step in mapping the metazoan gene network.

23. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M: **Protein interaction mapping in *C. elegans* using proteins involved in vulval development**. *Science* 2000, **287**:116-122.

24. Bouwmeester T, Bauch A, Ruffner H, Angrand PO, Bergamini G, Croughton K, Cruciat C, Eberhard D, Gagneur J, Ghidelli S *et al.*: **A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway**. *Nat Cell Biol* 2004, **6**:97-105.

25. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I *et al.*: **Transcriptional regulatory networks in *Saccharomyces cerevisiae***. *Science* 2002, **298**:799-804.

26. • von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P: **Comparative assessment of large-scale data sets of protein-protein interactions**. *Nature* 2002, **417**:399-403.
A critical assessment of yeast protein interaction data quality. Such tests will be important for animal protein interaction data.

27. Wu LF, Hughes TR, Davierwala AP, Robinson MD, Stoughton R, Altschuler SJ: **Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters**. *Nat Genet* 2002, **31**:255-265.

28. Hanisch D, Zien A, Zimmer R, Lengauer T: **Co-clustering of biological networks and gene expression data**. *Bioinformatics* 2002, **18(Suppl 1)**:S145-S154.

29. Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS: **A gene expression map for *Caenorhabditis elegans***. *Science* 2001, **293**:2087-2092.

30. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H *et al.*: **Systematic genetic analysis with ordered arrays of yeast deletion mutants**. *Science* 2001, **294**:2364-2368.

31. • Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M *et al.*: **Global mapping of the yeast genetic interaction network**. *Science* 2004, **303**:808-813.
First genome-scale analysis of synthetic lethal genetic interactions.

32. • Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK: **Global analysis of protein localization in budding yeast**. *Nature* 2003, **425**:686-691.
A *tour-de-force* mapping of subcellular localization of yeast proteins by GFP tagging and high-throughput microscopy.

33. Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, Jansen R, Liu Y *et al.*: **Subcellular localization of the yeast proteome**. *Genes Dev* 2002, **16**:707-719.

34. • Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M: **A Bayesian networks approach for predicting protein-protein interactions from genomic data**. *Science* 2003, **302**:449-453.
The authors integrate diverse yeast protein interaction data via a probabilistic approach, resulting in a more accurate integrated network.

35. Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J: **Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference**. *Nature* 2000, **408**:325-330.

36. • Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M *et al.*: **Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi**. *Nature* 2003, **421**:231-237.
The first large-scale targeted screen of loss-of-function mutant phenotypes in an animal, by systematic RNAi, identifying phenotypes for ~1700 mutants.

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

38. Walhout AJ, Reboul J, Shtanko O, Bertin N, Vaglio P, Ge H, Lee H, Doucette-Stamm L, Gunsalus KC, Schetter AJ *et al.*: **Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline**. *Curr Biol* 2002, **12**:1952-1958.

39. Lum L, Yao S, Mozer B, Rovescalli A, Von Kessler D, Nirenberg M, Beachy PA: **Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells**. *Science* 2003, **299**:2039-2045.

40. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO: **Protein function in the post-genomic era**. *Nature* 2000, **405**:823-826.

41. Galperin MY, Koonin EV: **Who's your neighbor? New computational approaches for functional genomics**. *Nat Biotechnol* 2000, **18**:609-613.

42. Huynen MA, Snel B, von Mering C, Bork P: **Function prediction and protein networks**. *Curr Opin Cell Biol* 2003, **15**:191-198.

43. Valencia A, Pazos F: **Computational methods for the prediction of protein interactions**. *Curr Opin Struct Biol* 2002, **12**:368-373.

44. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B: **STRING: a database of predicted functional associations between proteins**. *Nucleic Acids Res* 2003, **31**:258-261.

45. • Fraser AG, Marcotte EM: **A probabilistic view of gene function**. *Nat Genet* 2004, **36**:559-564.
An introduction to probabilistic gene networks, a logical step for functional genomics to take in defining the systems that make up a cell.

46. von Mering C, Zdobnov EM, Tsoka S, Ciccarelli FD, Pereira-Leal JB, Ouzounis CA, Bork P: **Genome evolution reveals biochemical**

networks and functional modules. *Proc Natl Acad Sci USA* 2003, **100**:15428-15433.

47. Date SV, Marcotte EM: **Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages**. *Nat Biotechnol* 2003, **21**:1055-1062.

48. Sulston JE, Schierenberg E, White JG, Thomson JN: **The embryonic cell lineage of the nematode *Caenorhabditis elegans***. *Dev Biol* 1983, **100**:64-119.

49. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G *et al.*: **A gene atlas of the mouse and human protein-encoding transcriptomes**. *Proc Natl Acad Sci USA* 2004, **101**:6062-6067.

50. Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlen M, Teeri TT, Lundeberg J *et al.*: **A transcriptional roadmap to wood formation**. *Proc Natl Acad Sci USA* 2001, **98**:14732-14737.

51. Andersson A, Keskitalo J, Sjodin A, Bhalerao R, Sterky F, Wissel K, Tandre K, Aspeborg H, Moyle R, Ohmiya Y *et al.*: **A transcriptional timetable of autumn senescence**. *Genome Biol* 2004, **5**:R24.

52. Kluger Y, Tuck DP, Chang JT, Nakayama Y, Poddar R, Kohya N, Lian Z, Ben Nasr A, Halaban HR, Krause DS *et al.*: **Lineage specificity of gene expression patterns**. *Proc Natl Acad Sci USA* 2004, **101**:6508-6513.

53. Reinke V, Gil IS, Ward S, Kazmer K: **Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans***. *Development* 2004, **131**:311-323.

54. Roy PJ, Stuart JM, Lund J, Kim SK: **Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans***. *Nature* 2002, **418**:975-979.

55. Wang J, Kim SK: **Global analysis of dauer gene expression in *Caenorhabditis elegans***. *Development* 2003, **130**:1621-1634.

56. Carson JP, Thaller C, Eichele G: **A transcriptome atlas of the mouse brain at cellular resolution**. *Curr Opin Neurobiol* 2002, **12**:562-565.

57. Pollet N, Delius H, Niehrs C: **In situ analysis of gene expression in *Xenopus* embryos**. *C R Biol* 2003, **326**:1011-1017.

58. Tomancak P, Beaton A, Weiszmann R, Kwan E, Shu S, Lewis SE, Richards S, Ashburner M, Hartenstein V, Celniker SE *et al.*: **Systematic determination of patterns of gene expression during Drosophila embryogenesis**. *Genome Biol* 2002, **3**:RESEARCH0088.

59. Gaudet J, Mango SE: **Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4**. *Science* 2002, **295**:821-825.

60. • Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP: **Gene expression during the life cycle of *Drosophila melanogaster***. *Science* 2002, **297**:2270-2275.
The authors illustrate the potential for monitoring developmental stage-specific transcriptional patterns.

61. Lu P, Nakorchevskiy A, Marcotte EM: **Expression deconvolution: a reinterpretation of DNA microarray data reveals dynamic changes in cell populations**. *Proc Natl Acad Sci USA* 2003, **100**:10370-10375.

62. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD *et al.*: **Functional discovery via a compendium of expression profiles**. *Cell* 2000, **102**:109-126.

63. Parsons AB, Brost RL, Ding H, Li Z, Zhang C, Sheikh B, Brown GW, Kane PM, Hughes TR, Boone C: **Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways**. *Nat Biotechnol* 2004, **22**:62-69.

64. Fay DS, Keenan S, Han M: **fzr-1 and lin-35/Rb function redundantly to control cell proliferation in *C. elegans* as revealed by a nonbiased synthetic screen**. *Genes Dev* 2002, **16**:503-517.

65. Einav Y, Shistik E, Shenfeld M, Simons AH, Melton DW, Canaani D: **Replication and episomal maintenance of Epstein-Barr virus-based vectors in mouse embryonal fibroblasts enable synthetic lethality screens**. *Mol Cancer Ther* 2003, **2**:1121-1128.

66. Simons AH, Dafni N, Dotan I, Oron Y, Canaani D: **Genetic synthetic lethality screen at the single gene level in cultured human cells**. *Nucleic Acids Res* 2001, **29**:E100.

67. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N: **Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data**. *Nat Genet* 2003, **34**:166-176.