techniques, as would point mutations in genes regulating growth, survival or differentiation. It will be important to apply higher resolution methodology, such as comparative genomic hybridization, to assess the genetic health of human ES cells during long-term culture, particularly as novel culture methods are developed and introduced into practice.

Another important lesson is that the karyotypically abnormal cell lines in this study still expressed canonical ES cell markers and retained the ability to differentiate *in vitro*, as indeed do certain cell lines derived from human embryonal carcinoma. Unlike the case of mouse cells, to which the bright-line criterion of germline chimerism can be applied, with human cells we lack robust biological markers that allow us to distinguish between normal ES cells and ES cells that have undergone minor genetic changes that might confer growth or survival advantages, or reduce their capacity for differentiation. Identification of such markers will be an important aspect of the evolution of standards for human ES cells.

Finally, no study to date has assessed the epigenetic stability of human ES cells. Changes in methylation status or in chromatin structure during culture *in vitro* may strongly affect the cells' capacity for differentiation. Though these issues may seem daunting, it is reassuring that mouse ES cells have been used throughout the world for over 20 years to produce normal live chimeric offspring, in the absence of any detailed or systematic assessment of the occurrence of spontaneous genetic or epigenetic alterations and their impact on developmental potential.

1. Draper J.S. *et al. Nat. Biotechnol.* **22**, 53–54 (2004).
2. Longo, L., Bygrave, A., Grosveld, F.G. & Pandolfi, P.P. *Transgenic Res.* **6**, 321–328 (1997).
3. Liu, X. *et al. Dev. Dyn.* **209**, 85–91 (1997).
4. Wilton, L. *Prenat. Diagn.* **22**, 512–518 (2002).
5. Thomson, J.A. *et al. Science* **282**, 1145–1147 (1998).
6. Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A. & Bongso, A. *Nat. Biotechnol.* **18**, 399–404 (2000).
7. Mitalipova, M. *et al. Stem Cells* **21**, 521–526 (2003).
8. Hovatta, O. *et al. Hum. Reprod.* **18**, 1404–1409 (2003).
9. Amit, M. *et al. Biol. Reprod.* **68**, 2150–2156 (2003).

Protein interaction maps on the fly

Peter Uetz & Michael J Pankratz

A protein interaction map for the fruit fly *Drosophila melanogaster* promises to facilitate functional analysis of many eukaryotic proteins.

Unraveling the ways in which proteins interact with each other on a genome-wide level is one of the main goals of proteomics research. Until now, large-scale protein interaction maps have been published only for yeast[1–3] and bacteria[4]. Now, in a paper published in *Science*, Giot *et al.*[5] report the largest protein interaction map to date and the first genome-wide study for a multicellular organism, the fly. Key to this work was the development of computational and statistical methodologies to extract relevant protein-protein interactions, which are presented in this issue[6].

Building the interaction map required a *tour de force* of PCR to amplify all 14,000 predicted *D. melanogaster* open reading frames (ORFs), of which more than 12,000 worked successfully. These PCR products were then cloned into two-hybrid bait (protein of interest) and prey (interactors) vectors, yielding roughly 11,000 clones each. This achievement is amazing by itself, given that sequencing of the PCR products represents a serious expressed sequence tag project on its own, which may allow biologists to identify many new exons and introns if investigated further. Next, all of the clones were screened against each other using methods Giot *et al.* have employed for the yeast genome. The fly project, however, proceeded on a vastly greater scale, requiring isolation of 45,417 two-hybrid positive colonies, from which 35,151 prey clones were obtained and sequenced. This approach generated 10,021 protein interactions involving 4,500 proteins.

Giot *et al.* then went one step further using their whole bait collection to screen a two-hybrid cDNA library prepared from embryonic, larval and adult tissues. Again, these screens resulted in another 45,962 positives of which 31,760 clones were successfully sequenced. This more traditional approach yielded 10,782 interactions involving 5,200 proteins. Thus, altogether Giot *et al.* churned out a whopping 20,405 interactions, dwarfing all previous efforts by at least four- to fivefold.

Nevertheless, two-hybrid screens are infamous for their false-positive rate; thus, one wonders about the reliability of these data. To address this issue, Giot *et al.* have developed fairly sophisticated statistics for the analysis of their protein network and Bader *et al.*[6] elaborate on this work, using yeast data as a model. In brief, both papers exploit topological and other information in a network to calculate confidence scores for each protein interaction. For example, two-hybrid interactions are considered low-confidence interactions if the interacting proteins are not found in a protein complex. Similarly, if proteins that have been found in a complex are far apart from each other in a two-hybrid map, the confidence score for this interaction decreases. Giot *et al.* use many more such parameters, including the reproducibility of interactions, the frequency with which bait and prey show up in a screen (indicating sticky and unspecific interactions) and evolutionary conservation of an interaction when compared with the yeast data set. After optimizing their statistical model, Bader *et al.* validated it by correlating their calculated confidence scores to independently collected information, such as gene ontology annotation (functional classification and subcellular localization) and gene expression data. As a result of such analyses, Giot *et al.* estimated that the fraction of 'biologically relevant' interactions may be as low as 11% in their whole data set, but around 38% in their high-confidence subset of 4,679 proteins and 4,780 interactions.

Given these quality-controlled protein interactions, what have we gained from this plethora of information? First, extrapolating from our experience with yeast, it is clear that although Giot *et al.* identified a huge number of interactions, they may represent only a fraction of all protein interactions in flies, which may harbor 50,000 or more. In addition, the number of essential interactions remains unresolved.

Second, the fly network has generated many novel hypotheses for biologists working on individual proteins, pathways or processes to investigate (see **Fig. 1**).

Third, 44% of all fly genes have homologs in the human genome, making *D. melanogaster* an excellent model system for studying normal and pathological processes; therefore, the protein interaction map should provide new avenues of investigation for medical research. In fact, Giot *et al.* explicitly highlight

Peter Uetz and Michael J. Pankratz are at the Institute of Genetics, Forschungszentrum Karlsruhe, Box 3640, D-76021 Karlsruhe, Germany.
e-mail: peter.uetz@itg.fzk.de

© 2004 Nature Publishing Group http://www.nature.com/naturebiotechnology

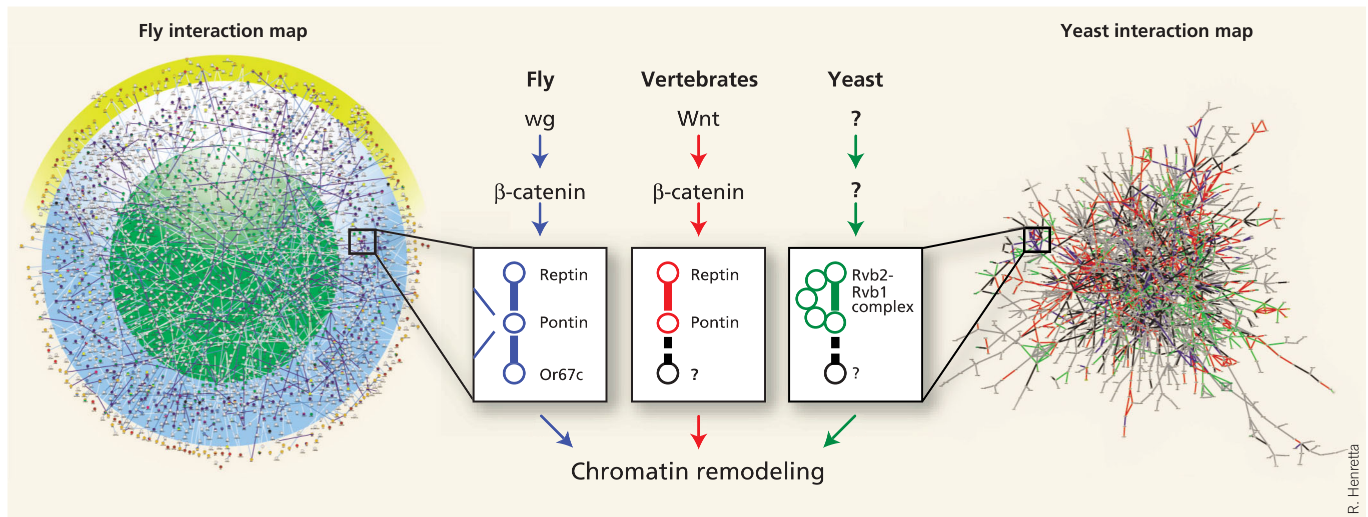

Figure 1 The awesome power of comparative interactomics. A comparison of interaction maps identifies evolutionarily conserved pathways in fly and yeast, such as those involving reptin and pontin that are homologs of DNA helicases. Reptin has just one high-confidence interactor[5], pontin (indicated by thick lines in the figure), whereas pontin returned two high-confidence proteins, one of which is reptin. Previous data showed reptin-pontin interactions in yeast, flies and vertebrates, and their role in chromatin remodeling[8], but those interactions and complexes involved different proteins in each model. In multicellular organisms, the pair is involved in wingless/β-catenin signaling. The other high-confidence partner of pontin, Or67c, is a membrane olfactory receptor. Although yeast does not have wingless/β-catenin homologs, it does have a membrane receptor for mating factor alpha, and mutants in reptin or pontin homologs show an impaired response to alpha factor[8]. Despite different interaction patterns in each model, comparative interactomics suggests both additional components in the pathway as well as the hypothesis that the ancestral role of reptin and pontin was in response to pheromone or chemical signals. Maps are from refs. 3 and 5.

© 2004 Nature Publishing Group http://www.nature.com/naturebiotechnology

disease-related proteins in their network, such as homologs of the oncoprotein Src.

Fourth, and most importantly, bioinformatic analysis will extract a wealth of new information from the fly and other protein networks. Comparing interaction networks of different model organisms is not only a powerful way to validate inherently unreliable data; it also allows systems biologists to identify conserved pathways and complexes[7]. Comparative analysis has already shown how proteins and their interactions evolve and, together with data from structural genomics, will tell us which residues in a three-dimensional structure are necessary for interactions and therefore function. In addition, interaction networks can be harnessed to predict protein interaction domains and motifs, and identify connections between pathways and processes that we would not have imagined otherwise. Topological analysis of protein networks also gives us an idea of why living systems are as robust to perturbations as they are and what complexity means in molecular terms.

Nevertheless, a great deal remains to be done. Most interactions have yet to be discovered, although we have only a few clues of how many there are. Two-hybrid interactions need to be verified by other methods, such as mass spectrometry analysis or systematic subcellular localization experiments. Interactions between proteins that are not localized to the same subcellular compartment or expressed in the same cell type and are not likely to be biologically relevant. Finally, for a full understanding of how a cell works, we need the structures of all components at atomic resolution. Structural genomics has a long way to go here.

But this wish list is mostly directed towards experimentalists. With more and more functional genomics data, bioinformatic analysis increasingly becomes a limiting factor. All these data sources have to be integrated so that a click on a protein in an interaction map pulls up more than a protein's structure or a multiple alignment of its relatives. Context-sensitive menus should pop up to give us a range of options to access a whole knowledge base of data. In fact, protein interaction maps are still static maps of squares and circles instead of realistic representations of what happens dynamically in a cell in four dimensions. So experimentalists and programmers pull up your sleeves and prepare for the exciting dawn of systems biology.

1. Uetz, P. *et al. Nature* **403**, 623–627 (2000).
2. Ito, T. et al. Proc. Natl. Acad. Sci. USA **98**, 4569–74 (2001).
3. Schwikowski, B., Uetz, P. & Fields, S. *Nat. Biotechnol.* **18**, 1257–1261 (2000).
4. Rain, J.C. *et al. Nature* **409**, 211–216 (2001).
5. Giot, L. *et al. Science* **302**, 1727–1736 (2003).
6. Bader, J.S., Chaudhuri, A. & Chant, J. *Nat. Biotechnol.* **22**, 78–85 (2004).
7. Kelley, B.P. *et al. Proc. Natl. Acad. Sci. USA* **100**, 11394–11399 (2003).
8. Jónsson, Z.O. *et al. J. Biol. Chem.* **276**, 16279–16288 (2001).