© 1997 Nature Publishing Group http://www.nature.com/naturegenetics

Looms to weave genomic nets

Roger Brent

Department of Molecular Biology, Massachussetts General Hospital and Department of Genetics, Harvard Medical School, 50 Blossom Street, Boston, Massachussetts 02114, USA

We are living in the middle of a time of transformation. It is sometimes easy to forget how fast the pace of change has been and to take time to consider observations that began as heterodox, but are now passing through the realm of novelty, on their way to becoming truisms.

Biologists are workers whose product is knowledge about living things—they use different techniques to generate information and add value to that information by analysing it. They then use the results of that analysis to guide further application of information-producing techniques. Typically, these have been performed by the same investigator in the same laboratory. We are now witnessing a migration of some of the production processes into facilities specialized for that purpose and find ourselves using new terms—for example, 'functional genomics—to describe aspects of biology that concern production.

Two main types of functional information are now produced industrially. One is gene-sequence data, generated, for example, by sequencing ESTs[2]. The other is data on the identity and abundance of mRNAs in a given cell population (sometimes referred to as transcript imaging) that can be obtained, for example, by hybridizing a population of mRNAs to gene sequences or oligonucleotides arrayed on glass or chips. Both kinds of information provide insight into function but neither allow immediate derivation of information that biologists, particularly geneticists, find particulary satisfying. That is, neither type of data allows biologists to position genes within pathways. A study by Micheline Fromont-Racine and collegues, presented on page 277 of this issue, marks a step towards the arrival of a third type of genomic information. Because it promises to aid the positioning of genes within pathways, it alters the biological enterprise as much as sequence information has already done.

Consider a single genetic pathway, for example, one that governs dosage compensation in the nematode worm (refs 4,5; Fig. 1*a*). In such a pathway, the named entities are genes, the arrows in it indicate which genes act on one another, and the nature of the arrowheads indicate positive or negative action. Information about this pathway was largely obtained by classical genetic and molecular manipulations. Fromont-Racine and colleagues now demonstrate another way to map connections between genes and their products. They have described a rapid approach to chart protein interactions—and hence, potential partners in genetic pathways—by multiple two-hybrid assays. They did this by working out a simple and reliable series of steps to deploy these assays to find interactors with proteins of interest, and then interactors with those interactors. The result is a set of techniques that can comprehensively chart connections among medium-sized (ten or so) sets of proteins involved in different processes.

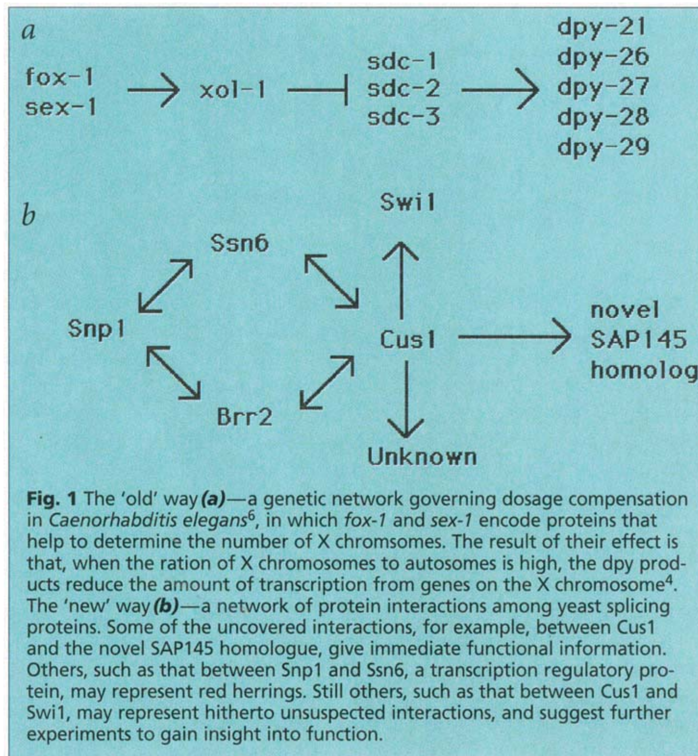

Fig. 1 The 'old' way **(*a*)**—a genetic network governing dosage compensation in *Caenorhabditis elegans*[6], in which *fox-1* and *sex-1* encode proteins that help to determine the number of X chromsomes. The result of their effect is that, when the ration of X chromosomes to autosomes is high, the dpy products reduce the amount of transcription from genes on the X chromosome[4]. The 'new' way **(*b*)**—a network of protein interactions among yeast splicing proteins. Some of the uncovered interactions, for example, between Cus1 and the novel SAP145 homologue, give immediate functional information. Others, such as that between Snp1 and Ssn6, a transcription regulatory protein, may represent red herrings. Still others, such as that between Cus1 and Swi1, may represent hitherto unsuspected interactions, and suggest further experiments to gain insight into function.

By way of example, the authors mapped connections between yeast proteins known to be involved in mRNA splicing. In doing this, they almost certainly isolated hitherto unknown proteins that function in this process and revealed unsuspected protein contacts that make biological sense and will guide future experiments. Not bad for a well-defined molecular mechanism in an organism whose genome is completely sequenced (see Fig. 1*b*).

A number of technical points contributed to their success. First, they used robust two-hybrid methods to detect interacting proteins. Second, they constructed high-quality libraries so that their chances of including all potential interactors were maximized. Third, they used interaction mating to speed the task of introducing potential interacting proteins into cells with their test proteins. Fourth, they established strong criteria for triage of candidate proteins, enabling them to deal with those interactors with a higher chance of biological significance. Fifth, they used some of the newly isolated proteins in baits for further screens.

The iterative strategy the authors describe is not the only route to construction of such protein maps. One approach is to introduce two whole-genome libraries into two different pools of yeast of opposite mating types, mate the pools, identify diploid cells that contained interacting proteins, and then sequence the genes encoding the proteins to identify the interactors[6]. Another is to mate ordered arrays of strains containing different baits with strains with different preys[7], a method that can be scaled up to analyse interactions among thousands of proteins. All of these approaches have two common features: their reliance on mature two-hybrid technology to increase the likelihood that interactions detected are real, and their use of yeast mating to vastly increase the number of interactions analysed.

The differences between interaction maps (or protein contigs) and classical maps of pathways are several fold. First, interaction maps are prone to 'noise', generated by irrelevant interactions with yeast proteins. They contain both false-positive interactions that are unlikely to be found in living cells, and

© 1997 Nature Publishing Group http://www.nature.com/naturegenetics

false-negative 'interactions'—interactions that would otherwise be detected but are not, perhaps because they are too weak to be detected in these particular reporter strains. Continued development of biological and informatic filters is likely to diminish the number of false positives, while the development of supplementary and robust physical techniques will increase the collecting power and reduce the number of false negatives.

Second, these maps are analogous but not identical. The arrows on the first kind of map indicate the action of genes on one another. The arrows on the second kind indicate contacts between proteins. As genes act through their protein products and much of the work that proteins do depends on their contact with other proteins, this information may be congruent with genetic information, but it may not. An extreme example is presented in Fig. 1; it is now known that Scd-2, Scd-3, Dpy-26, and Dpy-27 form a complex[4]. As Scd-2 and Sdc-3, and Dpy-26 and Dpy-27 act on the same stages of the genetic pathway, their physical associations were not foreshadowed by the epistasis experiments which yielded this information. Moreover, the arrows on the two-hybrid connection maps have two heads, that is, these experiments do not indicate which protein acts on the other.

Third, although the information contained in these experiments lacks the reach and power of classical genetic analysis, it can be used together with sequence and expression information to gain insight into gene function in organisms that are not easy to genetically manipulate. And, like sequence and expression data, 'connection' information can be produced industrially—rather than being practiced in seminar rooms and small labs, this information will be produced in factories.

So production of yet another type of genomic information stands on the brink of industrial scaling up. We don't know exactly how or where this data will be produced, although we can hope for anonymous high-tech facilities rather than dark satanic mills. With respect to sequence and expression data, we already have devices that could be considered the equivalent of spinning jennies; to weave this information into the fabric of functional knowledge inherent in pathway maps, we have hitherto used a variety of handlooms. The various mating approaches to acquisition of interaction information constitute the equivalent of the first Cartwright and Jacquard power looms. At the moment, these looms produce little cloth, and that cloth is of inferior quality. The day is coming, however, when these looms will be large enough to weave networks of functional information from the products of whole genomes.

1. Adams, M.D. *et al. Nature*, 337(supp) **3**, 17 (1995).
2. Lockhardt, D., et al., *Nature Biotech.* **14**, 1675–1680 (1996).
3. Avery, L. & Wasserman, S. *Trends Genet.* **8**, 312–316 (1992).
4. Meyer, B.J. *C. Elegans* II (Cold Spring Harbor Press, Cold Spring Harbor, New York, 1996).
5. Chuang, P.T. *et al. Science* **274**, 1736–1739 (1996).
6. Bartel, P.L. *et al. Nature Genet.* **12**, 72–78 (1996).
7. Finley, R.L. Jr. & Brent R. *Proc. Natl. Acad. Sci. USA* **91**, 12980–12984 (1994).

FISH with a twist

Paul M. Lizardi & David C. Ward

Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA

Biophysicists dream of a powerful microscope that would permit reading the sequence of DNA lying on a surface by observing chemical features of the individual bases, as if these were coloured beads on a chain. Meanwhile, cytogeneticists dream of detecting single-base changes inside cells by a glorified type of fluorescence *in situ* hybridization (FISH) using a standard fluorescence microscope. An article by Mats Nilsson and colleagues in this issue represents a significant step towards realizing this dream[1].

The study—resulting from a collaboration between Swedish and Danish scientists—is an application of multi-colour FISH with an interesting twist. The twist is in the use of a novel oligonucleotide probe that can be ligated to form a circle if it binds to a sequence of exact complementarity; in this study, the authors target a centromeric DNA repeat. The lateral 'arms' of the oligonucleotide twist around the DNA target, forming a double helix; their termini are designed to juxtapose so that they may be ligated enzymatically. A thermostable DNA ligase then converts the twisted probe into a circle, which, as a consequence of closure, becomes topologically trapped by its DNA target, like a hook around the tubing of a shower curtain. Immobilization of the probe upon closure led Nilsson and colleagues to coin the term 'padlock probe'[2].

Two different oligonucleotide probes were used in the current study, each corresponding to a different sequence variant of a centromeric alpha-satellite repeat. The sequences differ by only one base, and yet the DNA ligase used to circularize the probes is able to detect this difference, so that closure of the two alternative padlock probes occurs only in the event of perfect sequence recognition. Each probe sequence is designed so that it can be labelled with a unique fluorescent dye, allowing the two to be used simultaneously and revealed by the characteristic fluorescence emission of each dye. Until now, the detection of single base changes was outside the realm of what could be done with standard FISH technology. As the high specificity of thermostable DNA ligases is well-documented, there is every reason to expect that allele-specific ligation of padlock probes in FISH experiments will prove a robust technique.

Finding that FISH in the sea

The authors recorded images of fluorescence signals in the centromeric region of chromosomes 13 and 21 and discovered that the green and red signals display characteristic patterns that differ between individual chromosomes. The sequence variants are not chromosome-specific—any of the two sequence variants can be found in any of the two chromosome pairs. The chromosome-specific patterns can be identified by their unique combination of signal location and intensity. By analysing a population of ten individuals for these chromosome-specific patterns, they built a small pedigree that traces the origin of specific variants of chromosomes 13 and 21. Sometimes the padlock probe hybridization signals were observed as three distinct dots. The authors interpret these localized signals as evidence of differential chromatin condensation, but the data are insufficient to exclude other possible explanations for this phenomenon. The observed variability in signal intensity among different chromosomes 13 and 21 is interpreted as evidence of variations in the number of alphoid repeats.

The authors point out that variations in