MICROARRAYS

Clusters of clusters

The backlog of gene products without an assigned biochemical or genetic function has become an urgent problem for biologists. As the cost of sequencing continues to fall, and computational methods rapidly identify new genes, there is a great need to integrate gene sequences with the rest of biology. Bioinformaticians have made some headway in assigning predicted functions to new genes, primarily by using clustering algorithms, which assume that co-regulated genes have similar functions. These algorithms have limitations, however, that reflect the underlying parameters chosen by the investigator, and also the trade-off between an algorithm's accuracy and power. Now, Wu *et al.* describe a new approach to interpreting microarray data that eliminates some of the weaknesses of traditional clustering algorithms, facilitating large-scale predictions of *Saccharomyces cerevisiae* gene function.

Wu *et al.* tested their approach on a previously published microarray data set of 300 experiments, involving a variety of mutants and drug treatments of *S. cerevisiae*. Using several sets of parameters and algorithms, they generated partially overlapping transcriptional clusters. As each clustering algorithm defines similarity in a different way, combining several algorithms identified clusters that might have been missed by a single approach. An extra algorithm was added to search for co-regulated genes in a subset of experimental conditions, which again might have been missed by standard algorithms that build clusters based on global similarity in gene-expression profiles. Using statistical methods, the authors assigned a function to each cluster using the annotations in the Yeast Proteome Database (YPD) and the Munich Information Center for Protein Sequences (MIPS). Out of more than 13,000 derived clusters, 44% were assigned a function with a high level of confidence, arguing for the physiological relevance of these groupings.

To improve the predictive power of this approach, the authors removed particular algorithms that tended to generate large clusters, noting that small clusters were more apt to generate accurate predictions. They also added data from other experiments in the public domain, concentrating on areas of function in which the initial data set performed poorly — the cell cycle, signal transduction and differentiation. When applied to the larger, 424-experiment data set, this refined group of clusterings yielded predictions in 23 functional categories that were consistent with current annotation in at least 30% of the cases (a conservative estimate, with many giving a much higher rate of success).

All told, the authors assigned probable functions to ~1,650 poorly characterized yeast proteins, 285 of which are proposed to be involved in non-coding RNA metabolism. As always, biochemistry is the ultimate validation. Mutations in five genes that are predicted to be involved in ribosomal RNA processing confirmed their involvement in the accumulation or processing of different species of ribosomal RNA. As the authors point out, the flexibility of the approach indicates that there might be room for improvement, as well as for its application to the forthcoming data sets from higher eukaryotes.

Alan Packer, Associate Editor,
Nature Genetics

🌐 **References and links**
ORIGINAL RESEARCH PAPER Wu, L. F. *et al.* Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. *Nature Genet.* **31**, 255–265 (2002)
FURTHER READING Quackenbush, J. Computational analysis of microarray data. *Nature Rev. Genet.* **2**, 418–427 (2001)

ETHICS WATCH

Gene banks and blanket consent

Academic and industry researchers are understandably excited about the scientific and therapeutic potential presented by gene banks. Emerging areas of research, such as pharmacogenetics, require access to large pools of genetic data, which are of greatest scientific value when linked to identifiable personal information. For example, links between genetic information and health data are required to study gene–environment interactions effectively.

The ethical challenges that are associated with gene banks have attracted much attention from both academics and policy makers, as highlighted by the UK Human Genetics Commission's (1999) report. However, the issue of consent is particularly problematic for researchers. A strict application of many countries' consent laws requires researchers to obtain informed consent for the use of all identifiable genetic information and a fresh consent for new research projects, and to allow participants to withdraw their sample at anytime. Following these legal requirements is tremendously difficult for research involving large gene banks.

This dilemma has been dealt with through the increasing use of "blanket consent". Although one-time consents simplify the research process, the legal challenges associated with them have been underplayed. Because blanket consents are necessarily vague, they are too general to have much legal weight. Moreover, they do not allow patients to act meaningfully on their continuing right to control their health information. As such, most types of blanket consent fall "far short of true informed consent" (Greely 1999).

Given this legal uncertainty, adopting comprehensive blanket consent policies will require many countries to change their existing consent laws, requiring statutory amendments or legislation. However, creating new consent policy poses several significant challenges. First, there is some evidence that the public would not feel comfortable with a lowering of the legal standard to facilitate research. Indeed, a survey in 2001 of Canadians found that 90% strongly agree (61%) or agree (29%) that genetic information is different and that rules governing access to it should be stricter (Report for the BADMCC 2001). Second, jurisdictional variation in how this consent issue is addressed could add to the challenge of organizing large, multi-site research initiatives. In Canada, for example, the handling of health information is a provincial matter. As such, each of the ten provinces will need to craft legislation to address this topic.

Finally, governments need to consider carefully the broader social impact of changing consent law. Is it really time to jettison the well-established autonomy-driven view of consent? What additional safeguards will be needed to make such a change ethically sound?

Timothy Caulfield

REFERENCES Greely, H. T. Breaking the stalemate: a prospective regulatory framework for unforeseen research uses of human tissue samples and health information. *Wake Forest Law Rev.* **34**, 737–766 (1999) | Human Genetics Commission. *Whose hands on your genes?* [online] http://www.hgc.gov.uk/business_consultations 2maintext.pdf (1999) | Report for the Biotechnology Assistant Deputy Minister Coordinating Committee (BADMCC). *Public Opinion Research into Biotechnology Issues, Third Wave* (Pollara & Earnscliffe, Ottawa, Canada, 2001).