GENE FINDING

Seek and you shall find…

Genome biologists dream of defining a full complement of genes in a given genome, but gene prediction is notoriously difficult. Now, Kumar *et al.* describe a global strategy for gene finding that combines expression and homology data, with which they have discovered new *Saccharomyces cerevisiae* genes, and which is likely to be widely applicable.

Since the completion of the yeast genome sequence in 1996, ~60 new genes have been identified mainly from expression or homology data. Kumar *et al.* used these two information sources to carry out a global search of the yeast genome for previously unidentified genes. Their analysis began with a gene trap in which a promoterless, 5′ truncated *lacZ* reporter incorporated within a *Tn3* transposon was randomly inserted throughout the yeast genome. The resulting gene fusions were then detected in a high-throughput screen for β-galactosidase activity, and the trapped ORFs were identified by sequencing the insertion junctions. Previously unannotated ORFs identified in this way were studied further only if they were at least 25 codons long, in an intergenic region or orientated antisense to previously annotated genes. Their expression was then assessed by microarray analysis, in which labelled poly(A) RNA was hybridized to a specifically designed oligonucleotide array that carries sense and antisense sequence from the newly discovered ORFs. Only the ORFs that showed strong expression in these strand-specific assays were considered to be *bona fide* genes.

This experimental approach was also complemented with a computational analysis, in which Kumar *et al.* searched the yeast genome against metazoan and prokaryotic genome data and against the SWISS-PROT database.

Out of the total 137 new genes that Kumar *et al.* identified, 104 were less than 100 codons long — because short genes are notoriously difficult to predict, this is a reflection of the strength of this strategy for gene identification. The authors also uncovered a whole new family of subtelomerically located genes and augmented a class of previously poorly represented genes that overlap the already annotated genes but lie antisense to them.

FUNCTIONAL GENOMICS

…a complex web of interactions

As many captains of industry will tell you, people often work most effectively in closely interacting groups. The same goes for proteins, many of which function in multiprotein complexes — the workhorses of cellular life. However, such complexes are often large and dynamic, which makes them hard to study. Now, two teams report in *Nature* a high-throughput approach specifically tailored to analysing multiprotein complexes in *Saccharomyces cerevisiae*. Their findings provide a detailed view of the higher organization of this organism's proteome and provide valuable new information for annotating eukaryotic genomes.

Both teams took a similar approach to isolating protein complexes in *S. cerevisiae*. They began by adding an affinity tag to hundreds of yeast genes. When introduced into the yeast genome, these modified genes encode 'bait' proteins that can complex with other proteins under physiological conditions. Each bait can be captured via its tag, together with its associated proteins, which are then identified by mass spectrometry and computational analysis.

From a library of 1,548 tagged strains, Gavin *et al.* purified 589 bait proteins, from which they identified 1,440 distinct proteins that were present in 232 multiprotein complexes, 91% of which contained a novel component. Complexes were placed into functional categories based on literature- and database-derived information about the known proteins in each complex. Most complexes shared at least one component with another complex, forming a network of interactions. When organized by such connections, complexes from some functional categories grouped together, perhaps because their shared components reflect an underlying functional relationship. This analysis allowed Gavin *et al.* to propose roles for 231 of 304 proteins of unknown function. To assess whether the complexes that contain many highly conserved proteins are likely to be crucial to eukaryotic cells in general, Gavin *et al.* also purified proteins from three yeast multiprotein complexes — the Arp2–Arp3, Ccr4–Not1 and TRAPP complexes — from human cells, using the affinity tagging approach. They found the human complexes to be of a similar, if not identical, composition, indicating that this type of approach in yeast can inform studies of the human proteome.

Ho *et al.* used 725 proteins as baits, mostly kinases, phosphatases and components of the DNA damage response (DDR), from which they identified 1,578 proteins — 531 of unknown function — involved in 3,617 interactions. The authors analysed complexes associated with several important transcription factors and kinases, such as Fkh1, Kss1, Cdc28 and Dun1, a DDR component for which Ho *et al.* found many regulators and targets. Their use as bait of 86 proteins involved in the DDR allowed them to piece

© 2002 Macmillan Magazines Ltd

The total number of genes in yeast is unlikely to change as a result of this study — the number of previously predicted genes that turn out to be spurious is likely to be offset by the number of new predictions. The true impact of this study lies in the fact that it describes a method for re-examination of genome annotation that is applicable to other genomes and in its ability to predict the existence of genes that have so far eluded previously known methods.

Magdalena Skipper

References and links

ORIGINAL RESEARCH PAPER Kumar, A. *et al.* An integrated approach for finding overlooked genes in yeast. *Nature Biotechnol.* **20**, 58–63 (2002)

FURTHER READING Oliver, S. To-day, we have naming of parts… *Nature Biotechnol.* **20**, 27–28 (2002)

WEB SITE
Michael Snyder's lab:
http://www.yale.edu/snyder/res.html

together the network of interacting proteins that controls this response, so uncovering many new interactions of probable biological significance.

Although this approach is clearly very powerful, it is not without limitations — for example, Gavin *et al.* could not purify proteins under 15 kDa in size. Both groups also report a significant number of false-positive interactions, while failing to detect some known interactions, perhaps because the tag can interfere with a protein's function or with its physical associations. Although there is still a long way to go before we fully understand how a proteome's functional networks respond to the ever changing life of a cell, these two studies provide a panoramic view of protein function and a wealth of new functional data for genome annotation.

Jane Alfred

References and links

ORIGINAL RESEARCH PAPERS Gavin, A.-C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002) | Ho, Y. *et al.* Systematic analysis of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002)

FURTHER READING Kumar, A. & Snyder, M. Protein complexes take the bait. *Nature* **415**, 123–124 (2002)

WEB SITES
These data sets can be found at:
http://yeast.cellzome.com
http://www.mdsp.com/yeast

CANCER GENETICS

Intricate modelling

Lung cancer causes most cancer-related deaths worldwide, and of the different lung tumour types, adenocarcinoma is the most common. The involvement of *K-ras* in lung cancer was confirmed last year (see June 2001 Highlights), when Tyler Jacks' group showed that the somatic activation of a constitutively active *K-ras* allele (*K-ras*^{G12D}) can alone cause cancer. This study is now followed up by two new *K-ras*^{G12D} mouse models — one made by Jacks' group and the other by Harold Varmus' group — in which conditional gene-activation systems have been used to switch on this mutant *K-ras* allele. Importantly, these studies shed much needed light on the events required for lung tumour initiation, maintenance and regression, and are a step towards much better mouse models of cancer that can be used to develop and test new cancer therapies.

The gene-expression switches used by each team allowed them to ask different questions about lung tumour biology. Jacks' team used the Cre/*lox*P system to activate *K-ras*^{G12D} by targeting the endogenous *K-ras* locus with a 'lox–stop–lox' (LSL) *K-ras* allele in which *lox*P sites flank a transcriptional Stop element. When Cre was introduced into the lungs of LSL-*K-ras*^{G12D} mice through a nasally delivered adenovirus, their lungs became covered in precancerous lesions within four weeks, and there was evidence that Cre-induced activation of *K-ras*^{G12D} was responsible for this highly penetrant and rapid lung tumorigenesis.

However, such severe tumorigenesis is a problem — in last year's study, for example, the mice developed so many tumours that many died before the earlier lesions could progress to malignancy. The Jacks' group tackled this problem by lowering adenoviral-Cre doses to reduce tumour numbers, allowing the mice to survive and progress to later stages of tumorigenesis. Only then could the team solve the long-standing question of which of several early precancerous lesions give rise to adenocarcinomas — they report that it's most probably a lesion called atypical adenomatous hyperplasia, which seems to originate from one particular cell type, the alveolar type II cell. The authors also identified a new cell type, possibly a new lung stem cell, that might also contribute to adenocarcinoma development.

Fisher *et al.* used a different trick to turn on the *K-ras*^{G12D} allele — they created bi-transgenic mice that express both a tetracycline (Tet)-activatable form of *K-ras*^{G12D} and a reverse Tet transactivator protein expressed in alveolar type II cells that can only activate *K-ras*^{G12D} in the presence of doxycycline. This elegant approach allowed them to look at the events required for the initiation, maintenance and regression of lung tumours. Within one week of receiving doxycycline in their drinking water, these mice developed hyperplastic alveolar type II cells; after two months, their lungs became laden with large adenomas and adenocarcinomas. Nevertheless, within days of doxycycline withdrawal, these tumours regressed and underwent apoptosis. When these mice were crossed to two mouse strains, each null for a well-known tumour suppressor gene — *Trp53* or *p16*^{ink4a} — they developed more-aggressive tumours much more rapidly. Surprisingly, however, these tumours still underwent rapid apoptosis-mediated regression following doxycycline withdrawal, showing that this regression occurs via a p53-independent apoptotic pathway.

The fact that lung tumours with *p16*^{ink4a} and *Trp53* mutations can be induced to regress is good news indeed for those developing anticancer therapeutics, because *TP53* mutations are associated with tumour resistance to chemotherapy. But what is the pathway that mediates the p53-independent regression of these tumours? Fisher *et al.* have already found some clues to this question in their data, and to investigate it further, they plan to use microarray expression analysis to identify key transcriptional changes that occur during tumour induction and regression in these mice.

Jane Alfred

References and links

ORIGINAL RESEARCH PAPERS Jackson, E. L. *et al.* Analysis of lung tumour initiation and progression using conditional expression of oncogenic *k-ras*. *Genes Dev.* **15**, 3243–3248 (2002) | Fisher, G. H. *et al.* Induction and apoptotic regression of lung adenocarinomas by regulation of a *K-Ras* transgene in the presence and absence of tumour suppressor genes. *Genes Dev.* **15**, 3249–3262 (2002)

WEB SITES
Tyler Jacks' lab: http://mit.edu/biology/www/facultyareas/facresearch/jacks.shtml
Harold Varmus' lab: http://www.ski.edu/lab_homepage.cfm?lab=203

© 2002 Macmillan Magazines Ltd