happens in human tumours'. Concerned that their expression signature might be an artefact of their experimental cell culture system, the researchers analyzed the Global Cancer Map (GCM) database, which contains gene expression profiles from 190 primary human tumours of 14 different histological types [4]. They ordered the genes according to how closely their expression pattern approximated that of cyclin D1 and used the Kolmogorov-Smirnov (KS) nonparametric rank statistic [5] to capture the position of the set of 21 genes identified *in vitro* within this ordered list. This ranking is a similar process to that employed by Internet search engines. They found a significant correlation between the expression patterns of their set of target genes and the expression of cyclin D1 in human tumours. Further comparisons showed that cyclin D1 expression did not correlate with E2F target genes, but cyclin D3 expression did, highlighting differences between the cyclin D subtypes.

The rapid changes in gene expression indicated direct involvement between cyclin D1 and a transcription factor. C/EBPβ (CCAAT/enhancer-binding protein) alternatively named Nf-116 or LAP, was identified as a candidate. Further results obtained strongly supported the hypothesis that C/EBPβ is involved in regulating genes affected by cyclin D1 overexpression and an important player in previously unappreciated mechanisms of cyclin D1 action. Functional studies confirmed endogenous C/EBPβ as a crucial constitutive repressor of cyclin D1 target promoters and the existence of a functional interdependency between cyclin D1 and C/EBPβ for gene transcription, mediated by physical contact.

Controversially a study published simultaneously by Wang *et al.* identified peroxisome proliferator-activated receptor-γ (PPAR γ) as a target of cyclin D1 repression [6]. Using different methodology, their study compared wild type and cyclin D1 knockout mice.

Importance of independent activity

Earlier failures to correlate cdk-dependent cyclin D1 activity with tumour development has led to suggestions that cyclin D1 oncogenesis is cdk-independent [7]. Disruption of signalling through C/EBPβ is already known to contribute to malignant transformation in the murine mammary gland [8], thus this new work suggests cyclin D1 mediates its oncogenic effects via C/EBPβ. Nevertheless an involvement of PPAR γ in cancer is not unprecedented, as PPAR γ agonists have been shown to inhibit the growth of human colorectal cancer cells [9]. Thus, as stated by cell cycle researcher Richard Pestell, senior author of the Wang *et al.* study, 'the central issue is whether cyclin D1 inhibition of C/EBPβ or PPAR γ is the functionally relevant target'.

The research of Lamb *et al.* has wider implications because it highlights the role tumour gene expression databases could have in uncovering the functions of many other human oncogenes. Their methodology could become a prototype for future studies, as stated by Sridhar Ramaswamy, 'the paper represents a proof-of-concept of the approach'.

References

1 Lamb, J. *et al.* (2003) A mechanism of Cyclin D1 action encoded in the patterns of gene expression in human cancer. *Cell* 114, 323–334

2 Knockaert, M. *et al.* (2002) Pharmacological inhibitors of cyclin-dependent kinases. *Trends Pharmacol. Sci.* 23, 417–425

3 Coqueret, O. (2002) Linking cyclins to transcriptional control. *Gene* 299, 35–55

4 Ramaswamy, S. *et al.* (2001) Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. U. S. A.* 98, 15149–15154

5 Hollander, M. and Wolfe, D. (1999) Nonparametric Statistical Methods, (2nd edn), Wiley, New York

6 Wang, C. *et al.* (2003) Cyclin D1 repression of peroxisome proliferator-activated receptor γ expression and transactivation. *Mol. Cell. Biol.* 23, 6159–6173

7 Weinstat-Saslow, D. *et al.* (1995) Overexpression of cyclin D1 mRNA distinguishes invasive and *in situ* breast carcinomas from non-malignant lesions. *Nat. Med.* 1, 1257–1260

8 Zahnow, C.A. *et al.* (1997) Overexpression of C/EBPbeta-LIP, a naturally occurring, dominant-negative transcription factor, in human breast cancer. *J. Natl. Cancer Inst.* 89, 1887–1891

9 Sarraf, P. *et al.* (1999) Loss-of-function mutations in PPAR gamma associated with human colon cancer. *Mol. Cell* 3, 799–804

Minding the Ps and Qs of genomewide analysis

Henry Nicholls, BMN News

The much-loved p-value is inadequate for testing hypotheses on a genomewide scale, say statisticians, who argue that a new measure of statistical significance – the q-value – could help in these wide-ranging analyses.

Patterns for drug discovery

Mining genomes for significant patterns is one of the first steps in drug

 1359-6446/03/$ – see front matter PII: S1359-6446(03)02876-9

discovery. But sifting through the whole genome for correlations is like performing thousands of related experiments simultaneously, says John Storey of the Department of Biostatistics at the University of Washington in Seattle (http://www.washington.edu). It is a process that requires new statistical measures, he says.

One such measure, proposed by Storey, is the q-value, which assigns a value of significance to each of these 'experiments', but takes into account that thousands of them are being simultaneously considered.

This should be a useful tool for the drug discovery community, he says, and hopes that it 'motivates a closer look at the differences between p-values and q-values, and the implications of these differences on what we are willing to call valid discoveries,' he told *BioMedNet News* (http://news.bmn.com).

Ps and Qs

The well-known p-value measures significance in terms of the false positive rate – the rate at which null or uninteresting features are mistakenly called significant. But this approach is not tailored to the goals of genomewide analyses, note Storey and Robert Tibshirani in a recent publication [1].

By contrast, the q-value has emerged in answer to the needs posed by genomewide studies, and is a measure of significance in terms of the false discovery rate (FDR). The FDR describes the fraction of significant features that are expected to be false positives. 'For example,' said Storey, 'if I perform an experiment on every gene in the genome – say, via microarrays – and I select 100 genes to call significant, then a false discovery rate of 5% says that about 5 of these genes mistakenly made the list.'

This inversion of p-value logic should enable researchers to plumb the wealth of genomic data available, and make conclusions that strike a practical balance between excluding promising genes from and including erroneous genes in further analyses.

'I hope that the drug discovery community finds the q-value and false discovery rate concepts to be useful tools for mining these data in a statistically rigorous fashion,' said Storey.

Thorny statistical issues

Kim Zerba, a statistician at Bristol Myers Squibb in Princeton, NJ, USA (http://www.bms.com), agrees that the q-value should be useful in the exploratory stages of research. 'The q-value itself is not in widespread use in the industry, but people are familiar with the false discovery rate,' he said. 'It does add something to what we already have,' he told *BioMedNet News*.

However, the q-value is only a small step towards tackling the thorny statistical issues raised by the multiple testing that characterizes genomewide analysis, warns Zerba. 'This isn't going to solve the multiple testing problem,' he said.

Replication of a significant result, it seems, is still one of the most robust tools available. 'People are more focusing on replication,' said Zerba. 'But,' he admitted, 'that has got it's own issues as well.'

Reference

1 Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9440–9445

Infection stage clues to new antimalarial medicines

Vadim V. Demidov, freelance writer

Research of a key developmental phase of malarian parasite has revealed its 'Achilles heel' and could lead to robust antimalarial drugs and vaccines. This study involved the genome-wide timing of *Plasmodium falciparum* gene expression as the microbe asexually progresses through the intraerythrocytic cycle *in vitro* [1].

From complete DNA sequence to full transcriptome

Malaria was formerly linked to bad air (*mal aria* in Italian) near warm and humid swampy areas, where it is most common. But now this severe illness is associated with plasmodia, which are brought into human bloodstream by the bite of the female anopheles mosquito, the only malarian vector intrinsic to the tropical environment.

Recently, an international consortium has sequenced the genome of *P. falciparum* – the major cause of malaria [2]. Release and annotation of this genome sequence initiated large-scale proteomic–genomic studies of this protozoan to identify new therapeutic

1359-6446/03/$ – see front matter PII: S1359-6446(03)02881-2