of which affected the abundance of a staggering 35 proteins.

The authors took advantage of their previous transcriptomic studies in the same yeast population to compare the regulation of transcript and protein profiles. Despite some overlap between QTLs that control transcript and protein levels (including their arrangement in hotspots), there is little correspondence between the proteins and transcripts that are controlled by any one QTL.

The existence of regulatory hotspots is intriguing, as it agrees with our view of genetic disease as being caused by a single mutation that affects the expression of many proteins. Methods such as these that analyse protein profiles directly should therefore allow not only more efficient biomarker discovery but also a more accurate study of the role of mutations in development and disease.

Tanita Casci
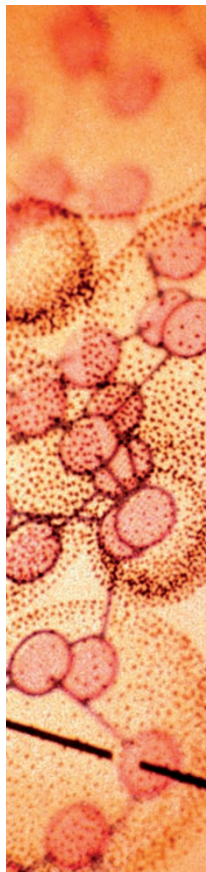
ORIGINAL RESEARCH PAPER Foss, E. J., Radulovic, D. *et al*. Genetic basis of proteome variation in yeast. *Nature Genet.* 21 October 2007 (doi:10.1038/ng.2007.22)

> Toxicity to the host [...] seems to inhibit gene transfer, and this toxicity might be attributable to gene dosage.

Of 61 loci that were uncloned from more than 5 genomes, half were present in single copies. These uncloned genes were mainly ribosomal structure and translation genes. When 39 of these uncloned genes were expressed in *E. coli* using an inducible expression vector, the *E. coli* cells died, indicating that the gene products were toxic. Computational analysis revealed that the genes that were tagged as unclonable in *E. coli* are predicted to be transfer-resistant throughout bacteria and archaea.

Toxicity to the host therefore seems to inhibit gene transfer, and this toxicity might be attributable to gene dosage. Understanding these limitations to HGT could have important implications for the choice of genes used to construct accurate phylogenetic trees.

Susan Jones, Senior Editor,
Nature Reviews Microbiology

ORIGINAL RESEARCH PAPER Sorek, R. *et al*. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 18 October 2007 (doi 10.1126/science. 1147112)

ASSOCIATION STUDIES

The best of the rest

The results of genome-wide association studies (GWAS) are commonly reported in terms of 'the top best' and a pool of other variants that didn't make the cut. A new approach promises to improve the ability of GWAS to detect disease mechanisms by considering not only the best-performing individual SNPs but groups of variants that belong to the same biological pathway.

The approach is borrowed from the analysis of microarrays, for which pathway-based methods are used routinely. Rather than scoring each gene expression change individually, here the data are searched for groups of genes that, when considered jointly, have significant differential expression compared with other sets of related genes. This approach has the advantage of detecting variants that individually make only a small contribution, and would therefore be missed from the conventional strategy of taking only the most significant hits. Extending this principle to GWAS means no longer settling for just the top-scoring 20–50 SNPs but instead considering groups of interacting genes.

Applying the analysis algorithm to GWAS required some technical adjustments to account for the differences between microarray and SNP data sets. For example, whereas on microarray chips one gene is represented by only a handful of transcripts, in GWAS each gene is represented by a variable number of randomly distributed SNPs, of which only one or a few are associated with the trait of interest. Given that each gene is associated with more than one SNP, how do you choose the most representative one? And how do you correct for the fact that longer genes have more SNPs than shorter ones? Although there is no straightforward answer, a useful strategy involves using the *P*-value of the most significant SNP to represent each gene. A more robust approach to condensing *P*-values for correlated SNPs to a single value to represent a gene should increase the power of pathway-based approaches.

The newly tailored algorithm as described by Wang *et al.* was tested on the GWAS data sets from two published studies of Parkinson disease and one of age-related macular degeneration. The SNPs were first matched to a gene; gene-annotation repositories were used to match gene identity to function and then, through gene-ontology databases, to a set of related genes. In both instances plausible biological pathways were identified, and in the case of Parkinson disease the approach implicated glycan-related genes in a potential new disease mechanism.

It should therefore be possible to move beyond simply looking at the most significant SNPs — lower-significance variants can be rehabilitated to further our understanding of complex diseases. Pathways are only as good as the annotations on which they are based, and so methods for incorporating previously neglected SNPs will continually improve, and with it hopefully the notoriously poor concordance among GWAS.

Tanita Casci

ORIGINAL RESEARCH PAPER Wang, K. *et al*. Pathway-based approaches for analysis of genomewide association studies. *Am. J. Hum. Genet.* 26 October 2007 (doi: 10.1086/522374)
**WEB SITE**
Maja Bucan's homepage: http://www.med.upenn.edu/camb/faculty/ggr/bucan.html