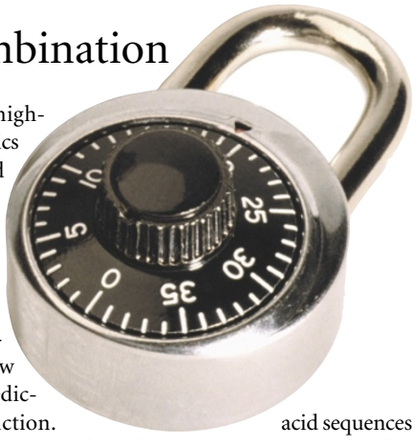
FUNCTIONAL GENOMICS

A cracking combination



As genome sequencing and high-throughput functional genomics approaches generate more and more data, researchers need new ways to tease out biologically relevant information. Two important advances in bioinformatics techniques — both of which rely on the combination of data sets — are now reported that allow better predictions of gene and protein function. Ge *et al.* have shown how gene-expression and protein–protein interaction data can be integrated to refine predictions of genetic interactions. And Dietmann and Holm have devised an improved computational method for detecting remote homology on the basis of structural similarities between proteins.

Ge *et al.* began their study by investigating whether genes that are co-expressed at the transcriptional level are more likely to encode proteins that interact with each other. They tested this in yeast, using transcriptional microarray data and a combination of genome-wide two-hybrid data, as well as individual protein data from literature searches. In general, clusters of co-expressed genes do indeed have a higher chance of giving a positive score in the two-hybrid assays and in experimental data derived from the literature. This lends weight to the idea that co-expression and protein interaction are both useful predictors of functional interaction. More importantly, the combination of these two types of data can also provide new insights into functionally related groups of genes. The authors provide an example in which a group of genes that is involved in a network of stress-response protein–protein interactions is divided into two co-expression clusters. This refines the view of how the genes interact and leads to the development of new, testable hypotheses.

Important clues about function can come from protein structure, especially if the proteins in question are distantly related and their amino-acid sequences have diverged. Dietmann and Holm combine sequence, structural and functional information to create a tree of protein structures, in which the most structurally similar proteins lie on the same branch of the tree. This approach overcomes some of the shortcomings of other bioinformatics tools. Their analysis can also infer homology relationships between proteins and group them into 'superfamilies'. The results compare very favourably with a protein classification system that is manually curated by experts (SCOP). Finally, the authors show that their automated system could be used to predict the superfamily membership, and therefore a putative function, of a new protein structure that might be determined as part of a structural genomics project.

Such computational approaches are essential if we are to predict and understand complex genetic networks from data generated by high-throughput genomic studies. And as the experimental technologies improve and generate more comprehensive and complementary data sets, the prospects for further combinatorial approaches to data analysis will be bright.

Magdalena Skipper

References and links

ORIGINAL RESEARCH PAPERS Ge, H. *et al.* Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nature Genet.* 10.1038/ng776 (2001) | Dietmann, S. & Holm, L. Identification of homology in protein structure classification. *Nature Struct. Biol.* **8**, 953–957 (2001)

FURTHER READING Lichtarge, O. Getting past appearances: the many-fold consequences of remote homology. *Nature Struct. Biol.* **8**, 918–920 (2001) | Brenner, S. E. A tour of structural genomics. *Nature Rev. Genet.* **2**, 801–809 (2001)

IN BRIEF

GENE EXPRESSION

Random monoallelic expression of three genes clustered within 60 kb of mouse *t* complex genomic DNA.

Sano, Y. *et al. Genome Res.* **15**, 1833–1841 (2001)

There are three mechanisms of gene-dosage control in mammals: random X inactivation, parent-of-origin autosomal gene imprinting and random autosomal inactivation. This last is very rare, but Sano *et al.* now report a cluster of three genes — *Nubp2*, *Igfals* and *Jsap1* — on mouse chromosome 17 that undergo this process. In single cells, these genes show X-like, random monoallelic expression, but can switch between active and inactive states during cell division — monoallelic expression correlates with the 50% methylation status of the genomic region. The authors discuss possible mechanisms for such gene expression patterns and their involvement in the biology of the *t* complex.

HUMAN GENETICS

An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing.

Pennacchio, L. A. *et al. Science* **294**, 169–173 (2001)

Apolipoproteins are known to affect plasma lipid levels in humans — an important factor in susceptibility to heart disease. Pennacchio *et al.* explored the already known apolipoprotein gene cluster on chromosome 11 for new susceptibility loci. By comparing mouse and human sequence, they identified a new locus, which when knocked out in mice leads to a substantial increase in plasma lipid levels, but when overexpressed causes these levels to drop below wild-type levels. This marked effect prompted the authors to look for SNPs in the human locus — all three rare alleles that they studied were associated with high lipid levels.

DEVELOPMENTAL BIOLOGY

Reciprocal mouse and human limb phenotypes caused by gain- and loss-of-function mutations affecting *Lmbr1*.

Clark, R. M. *et al. Genetics* **159**, 715–726 (2001)

Most of the dominantly inherited preaxial polydactly and syndactyly phenotypes, which affect limb-digit number, map to human chromosome 7q36 and to a homologous region in mice. By using deletion chromosomes, these authors show that limb defects that map to this region in mice result from gain-of-function mutations, rather than by haploinsufficiency. Mice with loss-of-function mutations in limb region 1 (*Lmbr1*), which might be allelic to the known limb morphology mutants *Hemimelic extra toes* and *Hammertoe,* have fewer limb digits than normal. Because this phenotype is reciprocal to polydactly, the authors propose that levels of *Lmbr1* activity control the morphology of the vertebrate limb skeleton.

© 2001 Macmillan Magazines Ltd