COMPARATIVE GENOMICS

Maximize your power

The power of comparative genomics has been amply demonstrated in recent years. So, the new report by Lander and colleagues, in which they compare four *Saccharomyces* species to identify genes and regulatory elements, might be taken as preaching to the converted — in reality, it shows the phenomenal power of comparative genomics in genome annotation.

To make the comparison, the authors sequenced *S. paradoxus*, *S. mikatae* and *S. bayanus* — species that are 5–20 million years apart — the widest distance being comparable to the sequence divergence between human and mouse. Sequence alignments showed extensive synteny, with most ORFs having a 1:1 correspondence and ambiguous alignments reflecting local gene family expansions that lie predominantly in telomeric regions.

To classify each ORF as biologically meaningful, the authors devised a reading-frame conservation test that was based on the proportion of the ORF that was conserved between the four species. Of the 5,945 ORFs tested, only 395 were not confirmed, most of which had no experimental support.

The four-way alignment also allowed much better prediction of gene structure, including start and stop codons — Kellis *et al.* found that the start and stop sites of some genes had been previously misannotated, which included some that pointed to sequencing errors.

Importantly, this approach also proved powerful for predicting regulatory motifs. To construct a comprehensive dictionary of such elements, the authors designed a *de novo* approach to discover regulatory sequences directly from the genome, and a way of inferring their function based on the knowledge about the function of genes with which they are associated. Kellis *et al.* also searched for pairs of motifs that occur in the same intergenic regions more often than would be expected by chance alone. This allowed them to begin to dissect the combinatorial control of gene regulation, providing information that might be used in future to construct motif interaction/regulatory networks.

As the authors point out, this comparative approach can be used for any set of species, and they suggest criteria for how such species should be selected for the best results. More complex genomes, such as our own, have a lower ratio of signal to noise; as a solution, the authors suggest filtering out repeated

sequences and increasing the number of genomes that are aligned.

Magdalena Skipper

References and links

ORIGINAL RESEARCH PAPER Kellis, M. *et al.* Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003)

HUMAN GENETICS

Cryptic ageing

The *LMNA* gene encodes two protein products, lamin A and lamin C — both are components of the nuclear lamina. Now, Eriksson *et al.* and de Sandre-Giovannoli *et al.* show that a silent mutation in this gene causes Hutchinson–Gilford Progeria Syndrome (HGPS), a rare disorder that results in premature ageing and shortened lifespan.

Eriksson *et al.* identified a 4.82-Mb region of chromosome 1q that was likely to contain the gene responsible for HGPS; but which of the ~80 candidate genes are involved? They started with *LMNA* because mutations in this gene have already been associated with a wide spectrum of human disorders. They sequenced this locus from 23 patients and discovered that 18 samples contained the same silent substitution in exon 11 (GGC>GGT). The authors were able to show that this mutation had occurred *de novo* in eight cases in which DNA samples were available from parents as well as sufferers. The lamin genes of the two patients sequenced in de Sandre-Giovannoli's study also had this mutation.

How could this silent substitution be responsible for such a severe phenotype? Reverse-transcriptase polymerase chain reaction and Western blot analysis showed that the mutations cause the partial activation of a cryptic splice site. So, although affected individuals express the wild-type lamin A, the mutant exon 11 allele produces a variant that lacks 50 amino acids of exon 11. This deletion might remove a proteolytic cleavage site that is needed to generate mature lamin A from its longer precursor. Lamin C is unaffected because it is only encoded by exons 1–10.

The high frequency of the silent substitution makes the development of molecular

diagnostics for this disorder feasible. But what is the genetic basis of the other five HGPS cases in Eriksson's study? One individual carried a conservative substitution at the same codon, which also activated the cryptic splice site, and another had a mutation in exon 2 of *LMNA*. However, fibroblast samples from uniparental-disomy and paternal-deletion carriers did not have any mutations in *LMNA*. The authors argue that mitotic crossovers might have occurred in the fibroblast cultures of these cells rescuing the mutant allele — however, they were unable to test this hypothesis.

Both groups went on to show that sufferers of HGPS have abnormal irregularly-shaped nuclear membranes. How this defect relates to premature ageing still remains a mystery. However, future studies using *Lmna*-knockout mice, and mice with a defect in lamin A processing — which have a phenotype resembling HGPS patients — should provide some answers.

Catherine Baxter

References and links

ORIGINAL RESEARCH PAPERS Eriksson, M. *et al.* Recurrent *de novo* point mutations in lamin A cause Hutchinson–Gilford Progeria Syndrome. *Nature* **25** April 2003 (10.1038/nature01629) | de Sandre-Giovannoli, A. *et al.* Lamin A truncation in Hutchinson–Gilford Progeria. *Science* **17** April 2003 (10.1126/science.1084125)