

HIGHLIGHTS

WEB WATCH

Human Protein Reference Database

• <http://www.hprd.org>
Browse, query or BLAST? Or why not view one of the available protein-interaction networks? It's up to you. The Human Protein Reference Database (HPRD) is a user-friendly resource that lets you do it all. However you find your protein of choice, information on every possible feature — alternative names, function, sequence, domains, motifs, interactions, expression, localization, post-translational modifications, substrates and any disease association — is just a click away.

Couple all of this with links to external resources, such as the OMIM, Swiss-Prot, LocusLink and Unigene web sites, and this really is a unified protein bioinformatics platform.

Almost all of the information that is included in the HPRD has been obtained manually by biologists who read hundreds of thousands of publications and interpreted and analysed the data. Every protein is reviewed twice, but should you spot any errors, they can be reported online. The HPRD ontology should soon be fully compliant with that of the Gene Ontology consortium, and the HPRD data will eventually be downloadable.

This resource is a joint venture between Akhilesh Pandey's laboratory (<http://pandeylab.bs.jhmi.edu>) and the Institute of Bioinformatics (<http://www.ibioinformatics.org>), but contributions from 'the outside world' are also encouraged — so how about becoming a 'molecule authority' for your favourite protein? And if your protein of interest cannot be found, then let the curators know and they will annotate it for you.

With more than 3,000 proteins in the HPRD already, and that number expected to reach 10,000 by the end of 2003, this really is a great resource for anyone who is interested in the human proteome!

Natalie Wilson



GENOME EVOLUTION

Silent garbage

For anybody who has turned their nose up at overflowing bins or discarded rubbish, the idea that junk could be made less intrusive is an attractive one. A new study shows that this is exactly the trick that a major class of genetic 'junk' — long interspersed elements (LINEs) — has managed.

The expression of LINEs is low in most cells, which means that for much of the time the host is not having its genome damaged by retroposition of these mobile elements or expending effort transcribing such 'selfish' DNA. Victoria Perepelitsa-Belancio and Prescott Deininger have now shown an important way in which expression of LINEs is kept low.

Their idea was that motifs in the LINEs' coding sequence could mimic the poly(A) signals at the end of an mRNA and therefore interrupt the expression of these elements. The first big hint that they might be correct came when they identified 19 potential poly(A) signals in the LINE-1 sense strand but only 2 in the antisense strand.

This was great circumstantial evidence that poly(A) signalling has a role in limiting LINE-1 expression but what was really needed was direct proof that the poly(A)s that they identified actually truncated LINE-1 expression. This proof came when they expressed a human LINE-1.3 element in a mouse cell line: they detected full length LINE-1.3 transcripts but these were much less abundant than smaller RNAs that corresponded in size to the species that would be expected if the internal poly(A) sites had truncated transcription. Equivalent experiments with a mouse

mobile element — LINE-1spa — produced qualitatively identical results, which indicates that this mechanism is conserved between the mammalian lineages that mice and humans represent.

Just how robust this system is in truncating LINE-1 expression was shown when the authors engineered mutations into the internal poly(A) site that corresponded to the most abundant transcript in their expression experiment. Predictably, the RNA band corresponding to truncation at this site was lost, but rather than leading to strong expression of the full-length transcripts, this led to more efficient use of other nearby poly(A) sites.

Perepelitsa-Belancio and Deininger also showed that endogenous transcripts that are truncated in the same way are present in a selection of human and mouse cells. Moreover, they showed that the 3' ends of most LINE-1 sequences in the human and mouse expressed-sequence-tag databases correspond to the positions of internal poly(A) sites.

So, it is clear that internal poly(A) sites truncate mobile element expression *in vivo* and therefore are one important way in which retroposition activity is limited in mammals. Perhaps one of the most interesting questions to arise from this work is how are these internal attenuation signals avoided in tissues or at developmental stages in which full-length transcripts are strongly expressed?

Nick Campbell

References and links

ORIGINAL RESEARCH PAPER

Perepelitsa-Belancio, V. & Deininger, P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nature Genet.* **35**, 363–366 (2003)

FURTHER READING Deininger, P., Moran, J. V., Batzer, M. A. & Kazazian, H. H. Mobile elements and mammalian genome evolution. *Curr. Opin. Genet. Dev.* **13**, 1–8 (2003)

WEB SITE

Prescott Deininger's laboratory: <http://129.81.225.52>