

Meeting report

Unscrambling the genome

Merlin Crossley

Address: School of Molecular and Microbial Biosciences, University of Sydney, New South Wales, 2006, Australia.
E-mail: M.Crossley@mmb.usyd.edu.au

Published: 30 April 2003

Genome Biology 2003, 4:320

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2003/4/5/320>

© 2003 BioMed Central Ltd

A report on the 24th Annual Lorne Conference on the Organization and Expression of the Genome, Lorne, Victoria, Australia, 16-20 February 2003.

The recent accumulation of sequence data has allowed more detailed analysis of both protein and RNA products, and their new roles in the function and evolution of organisms and organelles are being revealed. The Lorne Conference on the Organization and Expression of the Genome covered an interesting range of topics - from new functions of RNA to nuclear architecture and genome evolution.

RNA everywhere

Many people see proteins as more capable and versatile than RNA but recent work suggests new fundamental roles for this nucleic acid. Thomas Gingeras (Affymetrix, Santa Clara, USA) described an array study of the well-characterized human chromosomes 21 and 22 using closely spaced probes to detect expressed sequences. The results are astonishing: more than two thirds of transcripts detected on these chromosomes do not match any known gene and appear to be RNAs that do not encode protein. David Hume (University of Queensland, Brisbane, Australia), who has been working together with the Genome Exploration Research Group at the RIKEN Yokoyama Institute, Kanagawa, Japan, on sequencing full-length murine cDNAs, also reported that around a quarter of expressed mouse sequences appear to be non-coding RNAs. Although these RNAs are sometimes expressed at low levels, the sequence conservation between fish, mice and humans, and the finding that the transcription factors Sp1, Myc and p53 are often found bound to these RNAs' presumed promoters, suggests that the RNAs are genuine products of biological importance.

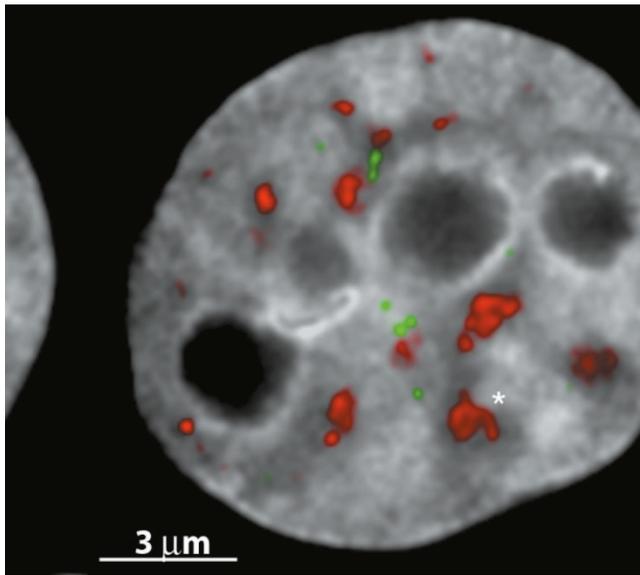
RNA - the long and the short of it

Although the functions of most non-coding RNAs remain to be defined, there are several examples for which significant

progress has been made. Denise Barlow (Institute of Molecular Biology, Salzburg, Austria) explained the role in parent-specific gene silencing (imprinting) of the mouse *Igf2r* cluster of a 108 kb long non-coding RNA termed *Air*. Truncating the *Air* RNA by inserting a polyadenylation and cleavage signal inhibits silencing of the *Igf2r* locus, suggesting that the RNA itself is critical. The *Air* RNA, which is encoded within the same gene cluster, is transcribed in the antisense direction relative to the *Igf2r* gene it silences, but it also turns off other genes within the cluster. Interestingly, it does not overlap with all the genes it affects, arguing against a simple antisense-mediated mechanism that is similar to RNA interference (RNAi).

Among the best understood non-coding RNAs are the microRNAs (miRNAs) that have been characterized by Victor Ambros and colleagues (Dartmouth Medical School, Hanover, USA). The first miRNAs that were identified, *lin-4* and *let-7*, are important developmental regulators in *Caenorhabditis elegans*. The *lin-4* miRNA can base-pair (albeit imperfectly) with seven sites in the 3' untranslated region of its target, the *lin-14* mRNA, and inhibits *lin-14* translation. The inhibition mechanism is unclear as *lin-4* neither causes *lin-14* RNA degradation nor prevents loading of the *lin-14* mRNA onto ribosomes. Because he noticed that known miRNAs are generated from 70 nucleotide hairpin precursors, Ambros searched the genomes of other organisms for sequences that can form such structures and tested the expression of candidates by northern blotting. Around 120 miRNAs have now been identified in the worm and a similar number in the genomes of higher organisms. The targets of these miRNAs, their mechanisms of action, and the steps in their synthesis are important issues now under investigation.

Short interfering RNAs, the intermediates in the RNAi pathway that mediates post-transcriptional gene silencing, have been used for several years to control gene expression in plants and animals artificially. Peter Waterhouse (Commonwealth Scientific and Industrial Research Organisation

**Figure 1**

The protein PSP1 marks a new type of nuclear body termed a paraspeckle. PSP1 is visible as bright spots and splicing speckles as dark spots (an example of a splicing speckle is indicated by the asterisk). Nuclear DNA is stained gray with diamidino phenylindole (DAPI). Image courtesy of Archa Fox, University of Dundee, UK.

(CSIRO) Plant Industry, Canberra, Australia) described his strategy for generating hairpin RNA that is effective in silencing homologous genes in plants. The inclusion of an intron that is spliced out to yield the short hairpin increases efficiency. Waterhouse described his new vectors pHANNIBAL and pHELLSGATE, which can be used in conjunction with recombinase systems *in vitro* to generate comprehensive libraries of silencing vectors. Steve Whyard (CSIRO Entomology, Canberra, Australia) described how hairpin RNA is being utilized in bio-control to combat insect pests and protect Australia from mollusc species introduced from other countries.

Stepping back to look at the nucleus

The first step in science often involves careful and detailed observations, and there have been several important new strategies employed for looking at nuclear substructures. Archa Fox (University of Dundee, UK) explained how purification and proteomic characterization of nucleolar components has provided insights into the function of the nucleolus. A large number of novel proteins were found in this compartment and further examination of one of these proteins, PSP1, marks a new nuclear structure termed the paraspeckle (Figure 1). PSP1 shuttles between the nucleolus and paraspeckles but when transcription is blocked it accumulates wholly within the nucleolus. Another nuclear body, the promyelocytic leukemia (PML) body, and its behavior in the face of heat shock has been investigated by David Bazett-Jones (Hospital for Sick

Children, Toronto, Canada). Most human cells contain 5-20 PML bodies, but these bodies are disrupted in promyelocytic leukemia. PML bodies contain the transcription factor PML, the transcriptional co-regulator CREB-binding protein (CBP) and other proteins, such as the small ubiquitin-like modifier (SUMO-1), but little DNA or RNA. It appears that CBP is positioned around the periphery of the bodies and the bodies are anchored by association with chromatin.

Organelles are a cell's best friends

The functions and evolution of conventional organelles, such as chloroplasts, are also becoming increasingly understood. Chloroplasts evolved from proteocyanobacteria. They contain DNA but many genes essential for chloroplast function appear to have moved from the chloroplast into the nuclear genome. How frequently this event occurs was estimated by Chun Huang (University of Adelaide, Australia). A nucleus-specific selectable marker gene (*neo*) was inserted into the chloroplast genome (plastome) of tobacco plants. Kanamycin-resistant plants, in which the marker gene had moved into the nuclear genome, were obtained at a frequency of 1 in 16,000, suggesting that DNA transfer occurs quite frequently. The work of William Martin (Heinrich-Heine Universität, Düsseldorf, Germany) supports the view that gene transfer from organelles to the nucleus is common. Martin compared the nuclear genome of *Arabidopsis* with three cyanobacterial genomes and 16 other prokaryotic genomes, in an attempt to estimate how many nuclear *Arabidopsis* genes originated from the ancestral chloroplasts. The data suggest that around 4,500, or about one fifth of all *Arabidopsis* genes, came from chloroplasts. Some, but not all, of these genes encode proteins that are targeted back into the chloroplast and are essential for its function.

The successful migration of a gene from a chloroplast to the nucleus requires not only movement of DNA sequence but also that suitable regulatory elements are present at the new location. But perhaps most interestingly, if the encoded protein is to be targeted back into the chloroplast then it will require the appropriate transit peptide. Geoffrey McFadden (University of Melbourne, Australia) has investigated the amino-terminal sequence extensions required to target proteins into the relict plastids (apicoplasts) of the malarial parasite. The characteristics of the transit peptide had been largely unknown, but McFadden's group has shown that the peptide must be rich in hydrophilic residues, particularly basic residues, and that binding sites for the chaperone protein Hsp70 (DnaK) are important. Bioinformatic searches were then used to identify a number of known and putative nuclearly encoded proteins that are targeted to the apicoplast. This information may help in building a picture of the biology of the malarial apicoplast, and because this plastid is required for the viability of the parasite, agents targeting its function may prove useful in the treatment or prevention of malaria.

In conclusion, the Lorne Genome Conference provides a venue at which data from different organisms are compared. The accumulating information demonstrates not only the complexity of the genome but also its dynamic nature, and takes us one step further towards unscrambling the puzzles of life.

comment

reviews

reports

deposited research

refereed research

interactions

information