

Meeting report

## Identifying transcribed sequences, and beyond

Ruth Wellenreuther and Stephanie Bechtel

Address: German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany.

Correspondence: Ruth Wellenreuther. E-mail: r.wellenreuther@dkfz.de

Published: 30 January 2003

*Genome Biology* 2003, 4:306

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2003/4/2/306>

© 2003 BioMed Central Ltd

---

A report on the 12th International Workshop 'Beyond the Identification of Transcribed Sequences (BITS): Functional, Expression and Evolutionary Analysis', Washington DC, USA, 25-28 October 2002.

---

Renaming the workshop 'Identification of Transcribed Sequences' to 'Beyond the Identification of Transcribed Sequences' a few years ago was perhaps a little premature, a workshop participant jocularly argued, so maybe it is now time to change the name back again. Indeed, after changing emphasis from genome to transcriptome and proteome analysis, the BITS workshop has returned somewhat to genes and transcripts in the context of 'regulonomics'.

### Identifying unusual transcripts

Despite the research now being directed towards functional analysis of transcribed sequences, identification of still unknown transcripts goes on. The huge size of the remaining task was indicated by talks and posters on mouse and zebrafish mutagenesis projects, large-scale cDNA sequencing, systematic screening for alternative splicing, and SAGE (serial analysis of gene expression) analysis. An example of the identification of 'unusual' transcripts was presented in the poster of John Conboy (Lawrence Berkeley National Laboratory, Berkeley, USA), who identified a class of mammalian genes with extremely distant 5' alternative first exons. The alternative first exons map up to 200 kb upstream of the coding exons and are most probably expressed from an independent promoter. These exons frequently include putative translation start signals in frame with the downstream coding region, such that the resulting alternative transcript codes for a protein with a distinct amino-terminal domain. Interestingly, a common feature among many of these genes is that they encode structural linking or 'adaptor' proteins within the cell. Among these

are: three genes for 4.1 protein, which anchor the cytoskeleton to the plasma membrane in red blood cells; three genes for ankyrins, which link the membrane skeleton to transmembrane proteins; three genes for cadherins, which have roles in muscle and neuronal cell association; and five genes encoding members of the MAGUK (membrane-associated guanylate kinases) family that act as scaffolding proteins in forming cell junctions.

### Functional analysis

Large-scale functional analysis of proteins encoded by the identified transcripts is under way, for example in mutagenesis projects, where functional information is inherent to the method in the form of phenotypes. Studying such phenotypes often reveals the pleiotropic effects of genes, as reported by Martin Hrabe de Angelis (GSF Research Center for Environment and Health, Neuherberg, Germany) in his talk on the large-scale mouse mutagenesis project of the ENU Mutagenesis Consortium, Munich. The disruption of one gene can lead to several different phenotypes; an example was provided by the Delta/Notch pathway, which is involved in neurogenesis. Hrabe de Angelis and coworkers found a novel *Delta1* gene allele in a mutant mouse line with defects in left-right symmetry.

Another large-scale functional analysis was presented by Michael Snyder (Yale University, New Haven, USA). Using arrays of recombinantly expressed and immobilized yeast proteins, Snyder and colleagues established binding assays for calmodulin, a calcium-binding protein affecting many enzymes, and the phospholipid signaling molecules phosphatidyl inositol-3-phosphate and phosphatidyl inositol-4,5-bisphosphate. A kinase activity assay has also been established in their lab. In their quest to screen for these diverse biochemical activities, they were able to find 33 new calmodulin binding proteins and derive a new calmodulin-binding site from analysis of them.

## Regulating expression

As transcript identification (including elucidation of unexpected phenomena in the transcriptome) and large-scale functional analysis continue, more emphasis is being placed on the regulation of gene expression at both the transcriptional and the translational level. The results of large-scale approaches to finding transcription factor binding sites or to measuring transcriptional or translational activity, as well as approaches to deciphering regulatory mechanisms, were presented at the meeting.

Yoav Arava and coworkers (Stanford University, Palo Alto, USA) analyzed the extent of ribosome association for thousands of yeast genes. They found lower ribosome density in the polysomes associated with large mRNAs, that is, there were fewer ribosomes per base of open-reading frame, suggesting lower translation efficiency for longer open reading frames. Arava and coworkers propose that the lower ribosome density of larger mRNAs may be a result of slower initiation rates, rather than low processivity of translation or slow termination rates, because they found similar ribosome densities on the 5' and 3' halves of the large mRNAs they studied. Gene regulation may also be the function of another poorly understood phenomenon, namely RNA editing. Daniel Morse (United States Naval Academy, Annapolis, USA) and colleagues looked for substrates for adenosine deaminases acting on RNA (ADARs) in human brain. Interestingly, editing in the 19 new human substrates they found was exclusively in non-coding regions. Yet another regulative issue was covered by Stefan Stamm (University of Erlangen-Nuernberg, Erlangen, Germany), who provided examples of the regulation of alternative splicing by protein phosphorylation induced by external signals. One example is provided by the splicing factor SLM-2 (SAM68-like molecule), which is phosphorylated at specific sites by tyrosine kinases. By mass-spectrometry Stamm and colleagues could identify three phosphorylated tyrosine residues. One of these tyrosines seems to regulate the intracellular localization of SLM-2, because knocking out the site by mutation leads to relocalization of the splicing factor from the nucleus to the cytoplasm.

Two mechanisms for coordinated regulation of gene expression acting at different stages of gene expression were presented: one that operates through mRNA stability and the other through coordinated translation. Arvind Raghavan (University of Minnesota, Minneapolis, USA) analyzed global mRNA decay in T cells. Using oligonucleotide arrays, the decay of mRNAs for approximately 6,000 genes was profiled in either stimulated or unstimulated T cells. Many transcripts encoding critical components of the T-cell-receptor pathway were coordinately downregulated following T-cell stimulation. About 50% of the induced transcripts have a short half-life of less than 90 minutes. Although numerous transcripts with rapid turnover rates contain sequences resembling AU-rich elements (AREs), many transcripts regulated at the level of mRNA stability do not contain any previously characterized

stability determinants. Raghavan and coworkers therefore plan to cluster the mRNAs according to their turnover characteristics and analyze their 5' and 3' untranslated regions, to seek new stability determinants. A model for the coordinated regulation of functionally related genes at the posttranscriptional level was proposed by Scott Tenenbaum and colleagues (Duke University Medical Center, Durham, USA), whose 'ribonomic profiling' revealed physical clustering of mRNA subpopulations by specific mRNA-binding proteins that recognize sequence elements common to the clustered transcripts. The steps of ribonomic profiling comprise isolation of mRNP complexes, isolation of their mRNA and protein components, and identification of the mRNA subsets within these complexes. From these profiles, Tenenbaum and colleagues discovered that mRNA-binding proteins are associated with unique subpopulations of messages, that the composition of the mRNA subsets can vary with cellular conditions, and that the same mRNA species can be found in multiple mRNP complexes. They propose a model in which functionally related genes are regulated post-transcriptionally as sub-populations by specific mRNA-binding proteins. By these means, some of the regulatory advantages of bacterial polycistronic mRNAs could be utilized by monocistronic mRNAs.

## Functional annotation

Several talks addressed the issue of functional annotation, for which integration of a variety of data from different sources will be crucial. Takashi Gojobori (National Institute of Genetics, Mishima, Japan) reported on an international annotation jamboree (The human full-length cDNA annotation invitational or 'H-invitational') that took place in Tokyo in late August. In this jamboree, 118 scientists compiled and interpreted data for functional annotation of more than 23,000 human cDNA clusters. Martin Ringwald (The Jackson Laboratory, Bar Harbor, USA) introduced the Jackson Laboratory mouse gene expression database, GXD [<http://www.informatics.jax.org/>], the scope of which is to integrate different kinds of expression data by standardized description of tissue of origin, expression pattern and experimental design. We (R.W.) showed details of LIFEdb [<http://www.dkfz.de/LIFEdb>], a database from which open reading frame (ORF) and intracellular protein localization data and information on novel human cDNAs can be retrieved. This database is currently being expanded to also provide information on protein expression and functional assays of these cDNAs.

In some respects we are beyond the identification of transcribed sequences, because functional analysis is well on its way, and the regulation of gene expression is coming into focus. But we are also still right in the middle of this process because of the underestimated importance and number of splice variants, as well as transcripts from genes with unusual structures, promise new adventures in transcript identification. Abstracts from the 2002 BITS workshop are available online at [<http://www.ornl.gov/meetings/bits2002/abstracts/toc.html>].