

GENOMICS

Gene expression in the worm

Large-scale analysis of gene expression in the nematode *Caenorhabditis elegans* takes advantage of its transparency and completely described cellular lineage.

An important aspect of gene function is when and where a gene is expressed. Global expression analyses are now being carried out in several systems, but a complete expression profile over the course of development has not as yet been achieved for a multicellular organism. Toward this goal, Don Moerman at the University of British Columbia, David Baillie at Simon Fraser University, and colleagues, now report an expression dataset for 10% of the genes in *C. elegans*, at all stages of development.

Moerman and colleagues were working on serial analysis of gene expression (SAGE) with purified cell populations in *C. elegans*. “We realized that SAGE has huge advantages for monitoring thousands of genes simultaneously,” he explains, “but you really don’t get the detail at every individual cell that you need for each gene.” The researchers decided to take advantage of the fact that the *C. elegans* cell lineage has been mapped, that is, that the identity of each cell in the organism at every developmental stage is known, and conduct a large-scale expression analysis using GFP reporters. This was greatly expedited by the decision to use overlapping PCR to generate fusions of promoter regions to GFP instead of standard cloning procedures and by the use of automated software for primer design.

“Making the DNA to generate transgenic worms was extremely rapid,” says Moerman. “And even making the transgenics was not as bad as one would think. The real slow step was the analysis of the expression patterns.” The researchers observed expression for 1,886 out of 2,402 genes, classifying them based on the pattern, and conducted high-resolution imaging on those with complex expression patterns. They compared the results with those in literature for 10% of genes,

and the data agreed in most (>80%) cases. Moerman emphasizes that one of the most important features of such a large-scale dataset is that it be very well validated. “These datasets aren’t any good if they’re not extremely accurate,” he says. “There is not a single cell where, if we say it expresses a certain gene, we are not 100% sure. When there is any doubt, it is better to say so.”

Accurate annotation is particularly important for perhaps the most exciting application of these data—the possibility to identify regulatory sequences that control gene expression in a particular cell or tissue type. “Although we cover only about 5% of the worm genome,” Moerman points out, “we have this rich array of expression patterns. If I was a bioinformatician interested in expression, I’d be all over this dataset.” The researchers saw tissue-specific expression for 20% of the genes analyzed, with many others expressed in multiple tissues or even in subsets of multiple tissues.

Perhaps the most striking result is that cell-specific gene expression, or expression restricted to a subset of cells within a single tissue, was very infrequent. This is consistent with previous SAGE data, where approximately a quarter of the genome was seen to be expressed in a given *C. elegans* cell type. But this also brings up a limitation of reporter studies such as this one. As Moerman puts it, “quantitative levels of expression, and what kinds of networks are formed, are going to be really important, and we’re barely scratching the surface of that. The qualitative on-off is sort of important to get you a first approximation, but then you have to look at the fine-tuning. There’s no reason that gene regulation should be as simple as we’d like it to be.”

Natalie de Souza

RESEARCH PAPERS

Hunt-Newbury, R. *et al.* High-throughput in vivo analysis of gene expression in *Caenorhabditis elegans*. *PLoS Biol.* **5**, 1981–1997 (2007).