

Comment

Funky, not junky

Gregory A Petsko

Address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.
E-mail: petsko@brandeis.edu

Published: 29 January 2003

Genome Biology 2003, 4:104

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2003/4/2/104>

© 2003 BioMed Central Ltd

As I write this, the holiday season is upon us again. It tends to arrive, in my case anyway, with the suddenness - and sometimes the consequences - of a train wreck. But amid the carnage there are moments of relative calm (by which I mean periods when one's relatives are calm and one can slip off to get some work done or just relax), and this year I've been using such moments to catch up on my reading.

I've been perusing 'Of Mice and Men'. No, not the sad, brilliant, haunting short novel by John Steinbeck, nor the delightful, insightful poem by Robert Burns that contains the phrase Steinbeck borrowed for his title. The particular Of Mice and Men I've been engrossed in is much more recent, and a lot longer. It came out last month and it has about 2.5 billion words. I had thought I could skip anywhere from 95 to 99% of it, since advance reviews had predicted that all but about 30,000 paragraphs would be unimportant. But now it looks as though I may have to plow through the whole damn thing. I had thought, in other words, that the draft sequence of the mouse genome would be like one of those paperback potboilers - something by, say, Harold Robbins or Jacqueline Susann - that one can just skim without missing much. Instead, it seems to be more like Proust's *Remembrance of Things Past*.

We live, as I'm fond of pointing out in these pages, in the age of genomics, and these days it seems as though every week brings us the complete genome sequence of yet another important eukaryotic organism. Yeast is followed by worm followed by fly followed by human, *Arabidopsis*, *Fugu* - the news comes so thick and fast that it's easy to get blasé. But even in these jaded times, the draft sequence of the mouse genome was front-page news. That's a lot of fuss, considering that as a result we can now take pride that our closest relative in terms of complete sequence information is a creature best known as a coward and a household pest. But I think all the publicity was justified. I would go even farther: I think the mouse genome sequence is more important, short-term, than the human genome sequence.

Most published commentaries on the mouse genome sequence, which was published on 5 December (*Nature* 2002, 420:520-562), have stressed the insights that will come from comparing it with that of *Homo sapiens*. That comparison has already started. For one thing, the mouse genome sequence validates the estimate of about 30,000 genes in the human genome, because *Mus musculus* appears to have about that number, and 99% of mouse genes have a human counterpart and vice versa. Even though the mouse genome contains 14% fewer base pairs (2.5 billion compared to 2.9 billion), over 90% of the two genomes can be partitioned into corresponding regions of conserved synteny (segments in which the gene identity and order in the most recent common ancestor has been conserved in both species). It is perhaps unsurprising that it is not the presence of unique set of genes that makes us human, but rather the way that a generic mammalian gene set is regulated.

At the nucleotide level, almost half of the human genome can be aligned to that of mouse, and the proportion of small (around 100 base-pair) segments in both mammalian genomes that is under selection can be estimated at about 5%, a number much higher than can be accounted for by protein-coding sequences alone. The implication is that in both mice and humans, the genome contains many 'non-coding' regions that are under selection pressure, and therefore that have important functions.

In many respects, the most intriguing part of the mouse genome sequence is what it has told us about these so-called 'junk' DNA sequences. In a mammalian genome, much of it originates from retrotransposons, which accumulate by being copied over and over into new locations. Given that retrotransposons reproduce via an RNA intermediate, they require reverse transcriptase to make the DNA copy that hops back into the genome. About 40% of the mouse genome seems to have derived from retrotransposons - about the same proportion as in the human genome. But in our genome the activity of these genetic parasites seems to be

low, with only about 100 still active. In mouse, by contrast, about 3,000 are actively jumping around. They come in three flavors: long interspersed elements (thousands of base pairs; 20% of the mouse genome) and long terminal repeats (10%), both of which produce their own reverse transcriptases, and short interspersed elements (about 300 base pairs; only about 8%), which do not.

What do these elements do? It appears they can do a number of things. When they jump into a host gene, they can disrupt its function completely or alter it subtly. They may be an important part of the clay that Nature uses to generate new functions as higher organisms evolve. Mart Speek has shown that long interspersed elements can alter the expression of neighboring host genes (see, for example, *Mol Cell Biol* 2001, **21**:1973-1985), indicating that they may be co-opted into becoming part of the machinery used to regulate transcription. John Moran and Jef Boeke have also shown that, in cultured human cells, they can cause deletion of large blocks of DNA (see *Cell* 1996, **87**:917-927). And Dixie Mager has suggested that regions of the genome free from accumulated retrotransposon sequences may be hallmarks of functional importance (*Genome Res* 2002, **12**:1483-1495), like the *Hox* genes that specify body plan.

How much of the 'junk' DNA will actually turn out to be useful for the cell is anyone's guess at present, but it seems clear that the term is outmoded. Calling something junk just because we don't understand what it does strikes me as narrow-minded. I suggest replacing this designation with 'funk' - functionally unknown DNA. I would prefer to think of our genome as funky rather than junky.

Mouse knockouts will certainly be important as disease models, and along with work in other model organisms mouse genetics and cell biology will be very useful in establishing what many mammalian genes are doing. And there's a school of thought that says that even though it's expensive to run a mouse lab, many biologists will now switch to doing mouse experiments because of how similar the mouse genome is to the human. But I'm not sure that the really interesting aspect of the mouse genome sequence will be what it can tell us about the location and likely functions of genes in the human genome. What has always fascinated me about the mouse is how easy it is to cure disease in mice and how frequently those cures fail to work in humans. Far be it from me to tell anyone what to do - OK, maybe not very far - but I think that the really interesting thing to figure out would be why, for example, so many treatments kill tumors in mice and don't do so in people. If I were the funding agencies, I'd be thinking about encouraging research that looks into the differences between mice and men, not just work that focuses on their similarities. The real excitement about the genome sequence of the model organism of the moment might just be the opportunity it affords to learn why a great model for understanding disease is so often a poor model for doing anything about it.

While we're waiting for the research directions to sort themselves out, we can keep busy trying to figure out what those couple of billion funky base pairs in the mouse, and human, genomes are doing, a process that will probably take decades. I suspect few of its founders would have thought, when they started the human genome project, that one of its consequences would be that thousands of scientists would spend their careers working on what used to be called junk. But then, the best laid schemes ...