

# Competition boosts bid to find human genes

Alison Abbott

The human genome sequence has perplexed researchers from the moment the draft version was assembled in 2001. The problem: our genome seems to contain remarkably few protein-coding genes.

The current estimate is between 20,000 and 25,000 — not many more than far simpler organisms such as nematode worms. But pinning down the exact number has proved to be a laborious business, and efforts have so far made only limited progress. Bioinformaticians meeting in Cambridge, UK, last week were optimistic that they can reverse this trend, thanks to a competition called E-GASP.

Launched earlier this year, E-GASP challenged 18 teams from around the world to develop better gene-prediction software for the human genome.

The initiative has had the desired effect of improving the available gene-prediction software, says co-organizer Roderic Guigó, a bioinformatician at the Municipal Institute of Medical Research in Barcelona, Spain.

Proving that a particular stretch of DNA is a gene involves doing an experiment to show that it is transcribed to make an RNA copy that can then guide protein production. But to do this for the whole genome would be time-consuming and expensive. Software that predicts the likely position of genes can speed things up, but often has only limited accuracy.

E-GASP aimed to improve matters using test material taken from 44 regions of the human genome — about 1% of its total length. For 13 of the regions, researchers at ENCODE, a US initiative to analyse all of the

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Researchers hope that advances in predictive software will speed the identification of genes.

functional elements in the human genome, painstakingly identified the position of all the genes by experiment.

This information was passed on to the 18 competing teams, who were then charged with predicting gene positions in the 31 remaining areas. At the same time, the ENCODE team completed its experimental analysis of the regions. Scientists gathered at the Wellcome Trust Sanger Institute on 6–7 May to hear the outcome.

“There was no absolute ‘right’ answer,” says Guigó. “Our annotation methods can only be described as ‘as-good-as-it-gets.’” So no overall winner was announced, although “a couple of the programs performed surprisingly well”, he adds.

Programs exploiting protein and transcription data provided the best predictions, but approaches involving comparisons with other genomes were also improved. Added together, the predictions put forward by the

competitors hit 70% of the genes identified by the ENCODE team almost perfectly.

Developing good prediction software is especially important for scientists working with species for which the genomes have been sequenced but little money is available for their analysis.

The new tools will also help guide the work of experimental scientists interested in human genes. The competitors’ predictions threw up hundreds of possible genes that weren’t identified in the lab experiments. ENCODE scientists in Barcelona and Geneva will select 200 of these for analysis in the next few months. “But based on our previous experience we do not expect more than 2% to be validated using our manual approach,” says Guigó.

He admits that other methods may turn up more genes. Researchers from the genomics company Affymetrix, based in Santa Clara, California, presented data to the Cambridge meeting from experiments using the latest generation of ‘microarrays’. These are made by chopping the genome up into thousands of bits of DNA and placing them, in order, on a grid. RNA will bind to the DNA it was produced from and therefore indicate any regions of the genome that are transcribed.

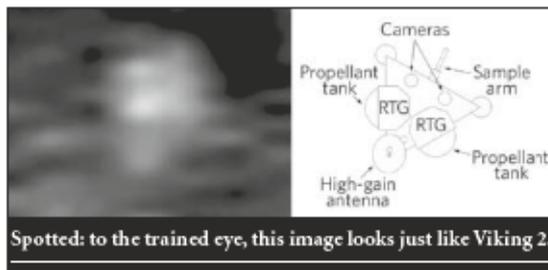
When the researchers washed RNA from a cell over the chip, 50% more regions on the grid bound to the RNA than there were known genes, suggesting that there is a lot more transcription going on than can be accounted for by genes identified so far. It isn’t known how much of this extra transcription represents new protein-coding genes, or whether some of the RNA molecules help to regulate existing genes. ■

## Early martian visitors are caught on camera

Tony Reichhardt, Washington

Scientists operating the Mars Orbiter Camera have spotted a pair of long-lost spacecraft — the Viking Lander 2 and Mars Polar Lander — in pictures taken from martian orbit. They hope that the pictures of Mars Polar Lander can provide clues to how and why the spacecraft crashed.

The Mars Orbiter Camera has been circling the planet since 1997. Researchers at Malin Space Science Systems of San Diego, California, who built the camera with the California Institute of Technology, are fairly sure about the identification of Viking 2, which has been sitting on the surface since 1976. But confirmation for the polar lander, which is thought to have crashed when its braking rockets shut down prematurely



Spotted: to the trained eye, this image looks just like Viking 2.

during its landing in 1999, will have to await sharper photos. Principal investigator Michael Malin hopes to begin taking those by late July, once frost on the martian surface has cleared up.

Malin has photographed other Mars landers from orbit, including Viking Lander 1, Mars Pathfinder and the two current Mars Exploration Rovers. But the panoramic

photos from Viking 2, which landed on a plain called Utopia Planitia, showed a flat and featureless terrain with few landmarks to help nail down where to look for it. Until now, the location was known only to within a few kilometres.

The picture believed to show the Mars Polar Lander was taken in 2000, although it was impossible to identify it at the time. Since then, photographs of the Mars rovers have given Malin a better idea of how parachutes and dark soil churned up by rocket blast would look from orbit.

So far, though, Malin hasn’t found anything in his pictures that looks like the small Beagle 2 lander lost in 2003. And until he has such a candidate, high-resolution searching of the surface would be like looking for a needle in a haystack. ■