

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

Mining the mouse

ArticleInfo		
ArticleID	:	4089
ArticleDOI	:	10.1186/gb-spotlight-20010518-02
ArticleCitationID	:	spotlight-20010518-02
ArticleSequenceNumber	:	160
ArticleCategory	:	Research news
ArticleFirstPage	:	1
ArticleLastPage	:	3
ArticleHistory	:	RegistrationDate : 2001-05-18 OnlineDate : 2001-05-18
ArticleCopyright	:	BioMed Central Ltd2001
ArticleGrants	:	
ArticleContext	:	130592211

Pete Moore

Email: pete.moore@dial.pipex.com

LONDON Although both [Celera](#) and the publicly funded [Mouse Sequencing Consortium](#) have [recently announced](#) draft sequences of the mouse genome, there is still a lot of work to do before a credible mouse genome is established.

On 27 April Celera stated that its whole-genome shotgun process had provided the company with a 6X coverage of the mouse genome, derived from three strains (129X1/SvJ, DBA/2J and A/J). Celera claims that its sequence covers more than 99% of the genome, with 95% in segments of at least 100,000 base pairs and 80% in segments of at least one million base pairs.

A solid achievement? Maybe not. Jane Rogers, head of the mouse genome sequencing effort at the [Sanger Centre](#) in Hinxton, UK was quick to question this announcement. "There is an issue with this. With what Celera have got, how do they know that they have the correct assembly?" Her concern is with the low marker density that is available to Celera. She can't see how the company's work isn't going to be heavily dependent on the data in the public domain, which Celera claims not to be using at the moment, although she knows that they have received computer tapes of the data.

Then, on 8 May, the Mouse Sequencing Consortium sent out a press release stating that its £40 million (\$58 million) project has now got 3X coverage of the sequence from one strain of mouse (C57BL/6J - commonly called Black 6). Its sequence covers 94% of the mouse genome.

The Mouse Sequencing Consortium, however, has not yet started to put its sequence together. It currently exists as 15 million individual unique sequence traces that, according to the consortium, are "small and unordered." "We are at the very beginning, we really need to finish the mouse genome because this will be a much more powerful tool than the draft," comments Steve Brown, head of the [MRC Mammalian Genetics Unit and UK Mouse Genome Centre, Harwell, UK](#).

If the proof of the pudding is in the eating, the proof of sequencing is in its ability to shed light on genes. Already the draft sequence is beginning to generate results.

"We and others have been able to use the draft sequence to identify some mouse genes we hadn't found before that are homologues to human genes that we know are involved in disorders," says Brown. "Given that the mouse is the pre-eminent model that we use to understand how humans function, we really need to know what genes are in the mouse genome. Then we can mutate and change them and look at the effects in mouse. So, identifying all the genes in the mouse that we know are present in the human is very important."

One example of this has been an announcement by Merck & Co that it has used Mouse Sequencing Consortium data to find the mouse equivalent of a human gene that may have a role in schizophrenia.

How about using mouse data to find new genes in the human genome? "We can compare the raw sequence in the mouse with the human sequence and look for areas of high degree of homology by just throwing the mouse sequence on top of the human one and scanning for those regions that are homologous. People are using this to discover new genes in both the mouse and human that simply

haven't been discovered before," Brown explains. This sort of cross-species comparison is becoming a powerful tool for filling out the annotation of not only the human genome but also the mouse genome.

One of the problems is knowing what has been found so far. As Rogers explains, researchers are unlikely to disclose partial findings. "If you mapped a human gene to within a megabase and you looked at what is coming out in terms of human annotation and couldn't find it, but you had a mouse read of it - wouldn't you keep quiet and go and look at the mouse?," she asks. "It will take a long time for the good results to come to light."

With the mouse genome in millions of pieces, it makes for a blunt research tool. "Clearly, we all hope that the money and energy will be available to go and finish the mouse sequence in the same way that we are going to have the human one," says Brown.

The Mouse Sequencing Consortium had a very limited task, to produce the 3X sequence. With that completed, the question is what happens next. "The next stage will be done through BAC [bacterial artificial chromosome] clones," says Rogers. "The [Genome Sequence Center](#) in Vancouver is in the process of constructing a database of fingerprinted BAC clones to provide a physical map resource around which to organize sequencing of the mouse genome. That is currently being tidied up by John McPherson at the [Washington University Genome Sequencing Center](#). From that, already there are clone contigs being selected to go into the sequencing pipeline. The majority of that work will be done at Washington and at the [Whitehead Institute](#). It's a methodology that proved successful for producing BAC-based physical maps of the *Arabidopsis thaliana* genome," she explains.

The hope is to complete the genome within a couple of years.

References

1. Celera, [<http://www.celera.com>]
2. Mouse genome page at the Sanger Centre, [http://www.sanger.ac.uk/Projects/M_musculus]
3. 'Mouse control' *Genome Biology* 10 May 2001 , [<http://www.genomebiology.com/spotlights/articles/SpotlightCompiler.asp?xml=20010510-3.xml&Status=Archive>]
4. Sanger Centre , [<http://www.sanger.ac.uk>]
5. Medical Research Council Mammalian Genetics Unit, Harwell, [<http://www.har.mrc.ac.uk/>]
6. Genome Sequence Center, Vancouver, [http://www.bcgsc.bc.ca/projects/mouse_mapping/]
7. Washington University Genome Sequencing Centre, [<http://www.genome.wustl.edu/gsc>]
8. Whitehead Institute, [<http://www.whitehead.mit.edu/home.html>]