

AUGUST 01, 2008

So Many Genes, So Little Time

“Deep Sequencing” Speeds Gene Discovery

Researchers have used a new "deep sequencing" technology to pinpoint a critical gene mutation in worms in only a few weeks. Deep sequencing allows scientists to simultaneously sequence millions of fragments of DNA at the same time. It would have taken years for scientists using older technology to sift through the worm's DNA until they found the proverbial needle in the haystack.

The researchers used the new technology, which has been likened to a genome center in a box, to identify a single gene mutation in the roundworm *C. elegans*. But the same method can be applied to identify gene mutations in any organism, including humans. The whole-genome analysis and gene identification took about a month and cost about \$5,000. It would have taken years of work and perhaps \$100,000 in salary and research costs using traditional analysis, said senior author Oliver Hobert, a Howard Hughes Medical Institute investigator at Columbia University.

Hobert and colleagues reported their findings on August 1, 2008, in an advance online publication in the journal *Nature Methods*.

"As the sequencing capacity of the technology increases, whole-genome sequencing will become the method of choice for higher organisms, including fruit flies, mice and humans."

- Oliver Hobert

Scientists have long used "forward genetic" approaches to explore gene function. A classical forward genetic analysis starts with a physical characteristic (called a phenotype) of interest and ends with the identification of the gene or genes that are responsible for that phenotype.

In a typical forward genetic strategy, researchers use chemicals or radiation to mutate genes at random in the cells of the plants or animals they are studying. After mutations are introduced, the scientists screen the mutants to identify those with a phenotype that is of interest. The changes in a cell or organism's phenotype are then attributed to the mutated genes and, by inference, to their protein products.

The next step is to try to pinpoint the gene mutation responsible for producing the defect observed in the mutant. However, finding the mutant gene can take years of laborious cross-breeding of mutant and normal plants or animals, with scientists carefully identifying the offspring that inherit the trait. At that point, researchers can begin using genetic markers to zero in on the region of the genome that carries the mutation. Once the region of the genome is narrowed, scientists use gene sequencing technology to pinpoint the mutation. Identifying mutations in humans involves tracing the inheritance of a disease or other trait in a family and determining the pattern of inheritance of genetic markers to narrow down the region of the genome that contains the mutation.

In contrast, the whole-genome-sequencing approach used by Hobert and his colleagues bypasses the need for cross-breeding or inheritance analysis. In the experiments reported in *Nature Methods*, they used a whole-genome-sequencing technology developed by Illumina, a company based in San Diego.

Hobert's team first used a chemical to produce a massive number of mutations in the genome of *C. elegans*. They identified one mutant worm that showed a particular defect in generating a specific type of sensory neuron. But instead of going through a long series of cross-breeding experiments to pinpoint the responsible gene, they used the Illumina technology to sequence the entire genome of the mutant worm in a general region they identified as relevant.

Hobert and his colleagues began by snipping the worm genome into millions of very short DNA segments, like chopping a masked picture into a huge number of puzzle pieces. The DNA segments were then attached to a transparent slide and massive numbers of copies of each DNA segment were created by the Illumina genome analyzer system. Then, each segment was sequenced using fluorescent-tagged chemicals, like unmasking each piece of the picture puzzle. Finally, sophisticated computer analysis, done by Hobert's colleague Itisk Pe'er at Columbia University, organized the mass of DNA sequence data on the DNA segments to assemble a map of the whole genome sequence -- like assembling the puzzle pieces to reveal the picture.

This sequencing revealed some 80 DNA sequence variations between the mutant worm and the known standard *C. elegans* genome sequence. Manual re-sequencing revealed that the majority of those variations were either sequencing errors, were due to random variation among individuals, or were not part of genes that code for functional proteins in the worm. Only four of

the true variations occurred in protein-coding regions of the genome, and the researchers readily narrowed down those variations to reveal the relevant gene mutation, said Hobert.

“In the short term, , this will become the standard method of mapping genes in *C. elegans*,” said Hobert. “And over the longer term, as the sequencing capacity of the technology increases, whole-genome sequencing will become the method of choice for higher organisms, including fruitflies, mice and humans.” Hobert also pointed out that other companies are developing different approaches to deep sequencing and those contributions will also drive progress. He said that scientists' ability to rapidly and cheaply pinpoint mutations will encourage far more extensive genetic screens.

“Forward genetic screening of mutants is conceptually extraordinarily elegant,” said Hobert. “In reality, however, people have a hard time fully and exhaustively harvesting the fruits of many genetic screens because it takes so much time to map and identify the molecular lesions in isolated mutants.

“Moreover, with deep sequencing technology, one can now focus gene characterization not just based on whether a gene has an interesting phenotype but also based on whether the molecular identity of the gene is of interest,” Hobert said. “For example, in the past, if you came up with fifty worm mutants that were unable to crawl, you had to more or less randomly pick which mutant to pursue. But with whole-genome, deep sequencing, at least in principle, you could sequence all fifty mutants and ask which genes look the most interesting in terms of the molecular identity - for example, because it has a human homolog - and further characterize those. As importantly, deep sequencing also allows you to identify the molecular identity of mutants that would have been otherwise very hard to map by conventional methods, such as behavioral mutants or modifier mutants. This whole technology is a true blessing for classic genetic analysis.”