Researchers develop complete map of mouse genetic variation

September 26 2011

For decades, laboratory mice have been widely used in research aimed at understanding which genes are involved in various illnesses. But actual variations in past gene sequences of mice were unknown. While researchers were able to determine that a variant affecting disease was in a certain region, they couldn't pinpoint the exact set of variants in that region.

Now, in new research recently published in the journal *Nature*, an international team of investigators that included UCLA researchers reports that it has sequenced the complete genomes of 17 strains of mice, including the most frequently used laboratory strains. The massive genetic catalog will provide scientists with unparalleled data for studying both how genetic variation affects phenotype and how mice evolved.

Researchers from UCLA's Henry Samueli School of Engineering and Applied Science played a key role in the study, using UCLA-developed technology to help sequence a nearly complete map of mouse genetic variation. Cataloging the full set of variants is a first step in identifying the actual variants affecting disease.

"The actual number of variants discovered is important because this gives the complete picture of how much variation exists in these [mouse strains](#)," said Eleazar Eskin, an associate professor of computer science at UCLA Engineering who develops techniques for solving computational problems that arise in the study of the [genetic basis](#) of disease. "Our group here at UCLA, and others, had tried to estimate this

number from the data that existed previously, which only collected a fraction of the total variation."

The new study was led by groups from the Wellcome Trust Sanger Institute and the Wellcome Trust Centre for Human Genetics in Oxford.

Previous technology used in genetic sequencing would, in some cases, make ambiguous predictions, and the locations of these ambiguities resulted in missing entries in the catalog of genetic variation in mice.

"Our role in the collaboration was to apply a technique that we developed a couple years ago for predicting variants where the sequencer failed to make a prediction," said Eskin, who holds a joint appointment in the department of human genetics at the David Geffen School of Medicine at UCLA. "Our technique, called imputation, uses the complete data to try to fill in some of these entries. The method, called EMINIM, was specifically designed for mouse data. Our contribution was to apply this technique to the data, which led to an increase in the number of variants identified."

With the full set of genetic information, researchers can now accurately predict the phylogeny — similar in concept to the family tree — of how the various mouse strains are related. The new study confirms that mice have a complex evolutionary history.

In addition, the study has some applications for the new genetic map, which were impossible before the development of this resource. One application involves identifying "allele specific expression." This expression describes the activity level of a gene. Each individual has two chromosomes, one from the mother and one from the father. For this reason, there are two copies of each gene.

Previous methods that would measure expression levels or activity levels

of genes would measure the combined activity level of both copies of the genes. The genetic map generated by this study now allows researchers to measure the activity level of each copy individually.

"What the study confirms," Eskin said, "is that for many genes, these expression levels differ quite dramatically. This type of analysis was very difficult to perform before such a study."

Provided by University of California - Los Angeles