The 1% solution

By Lauren Martz, Staff Writer

University of Washington researchers think they have found a shortcut to identifying rare genetic variants that contribute to inherited disorders. Instead of mining the entire genome, the team is focused on the exome—the protein-coding region of the genome that makes up about 1% of total genetic material. The theory is that many genetic disorders can be mapped to the exome because they result from mutated proteins.

The group now has published proof of concept for the approach in *Nature*¹ and thinks it could speed up discovery of rare disease markers and lower the cost of discovery by reducing the amount of genetic real estate that needs to be covered.

The team chose to study the rare, autosomal dominant mendelian disorder Freeman-Sheldon syndrome (FSS). The disease is characterized by joint contractures, muscle weakness and fibrosis, and it has a known causal variant: *MYH3* (*myosin heavy chain 3 skeletal muscle, embryonic*).

Researchers used the GA2 platform from **Illumina Inc.** to sequence the exomes of four FSS patients and eight individuals without the disease from the HapMap and Human Genome Structure Variation project.

In testing to see if the method could pick out the causal gene in the FSS patients, common genetic variants that were also present in the HapMap individuals or in published genetic databases were eliminated. A scan of the remaining mutations isolated *MYH3* as the only candidate gene mutated in all four FSS patients.

“What we did was use a commercial sequencing method in a new way,” said Jay Shendure, assistant professor of genome sciences at the University of Washington. “Here, we can focus on the 1% of the genome. Cost-wise it is cheaper to focus on the subset.”

Jorge Conde, CEO of **Knome Inc.**, told *SciBX* that “this is the first time I’ve ever seen a large-scale publication of exon sequencing looking at candidate genes for a particular condition. The method can quickly narrow down the gene hunt to one gene.”

Knome offers KnomeSELECT, which is an exome sequencing and analysis service, and KnomeCOMPLETE, for whole-genome sequencing.

Conde said the benefit of focusing on the exome “is that it provides more bang for your buck by concentrating on the regions that we know most about and that are most likely to harbor the mutations we seek. The exome is about 1% of the entire genome, and it holds 60%, 70% or maybe even 80% of the genetic information that we know about today. It provides a lot higher cost performance if we restrict searches to that area.”

“What we did was use a commercial sequencing method in a new way.”

—Jay Shendure,
University of Washington

According to the authors of the *Nature* article, 20 exomes could be sequenced for the cost of 1 whole genome.

Shendure told *SciBX* that “the next steps are to further optimize the technology, which could improve the cost equation.”

The limits of limited sequencing

Despite the clear cost advantage of exome sequencing versus whole-genome sequencing, there is potential for the former approach to miss genetic variants.

“The problem with execution of the *Nature* study is that the needle has been placed very carefully in the haystack,” said Bud Mishra, professor of computer sciences and mathematics at **New York University** and professor of cell biology at the NYU School of Medicine. “The team knew where the mutation should be and had substantial information about the target gene. In 2006, it was known that FSS is caused by a single nonsynonymous, monogenic mutation that exists in the *MYH3* gene of the exome.”

As a result, said Mishra, the authors “knew that through only reading exomes and filtering out common genes that the gene of interest would still be present. I would rather work on good quality, whole-genome sequencing. The benefit of cost and simplicity of exome sequencing is appealing but could be fooling. In haplotype and whole-genome sequencing, there is nothing hiding. You are not throwing out pieces of genetic information that you might need later.”

Clifford Reid, CEO of **Complete Genomics Inc.**, told *SciBX* that the exome strategy could miss noncoding regulatory sequences.

Complete Genomics offers a suite of genome-sequencing services that includes exome sequencing.

Shendure said his team has successfully applied the method to characterize the previously unknown genetic underpinnings of Miller syndrome, an autosomal recessive mendelian disorder characterized by extensive facial and limb defects.

The team is currently applying exome sequencing analysis to more diseases.

Conde noted that exome sequencing requires the additional step of exon capture, which consists of preselecting a genome’s protein coding regions using microarray-based hybridization. He said adding steps to genetic sequencing procedures makes them more complicated and can increase the chance of errors.

Mishra said it will be important to have a standard method to determine how many exomes need to be analyzed in a given study. “The researchers chose eight individuals from the HapMap project and four patients with FSS, but it is a problem that they did not show how they got this number. They do show, however, that if you shrink the number of patients from four to three, you get three genes as your final answer rather than one.”

Shendure told *SciBX* his team is “working on models that estimate the number of individuals that would be required under various assumptions

about the underlying genetics. The general rule will be the more complex and heterogeneous the disease, the more individuals will be required.”

He also noted that the more data on variation in normal individuals collected, the more powerful the approach would become.

Going forward, Shendure said, “the method can be applied to other mendelian diseases, and we are pushing it in the direction of complex diseases, which are analytically much more challenging.” To do so, he said, it will be necessary to scale-up the method to larger population sizes and develop better analytical tools.

Shendure told *SciBX* that the method has not been patented because it represents an application of existing hybridization-based methods.

Martz, L. *SciBX* 2(36); doi:10.1038/scibx.2009.1364

Published online Sept. 17, 2009

REFERENCES

1. Ng, S. *et al. Nature*; published online Aug. 16, 2009; doi:10.1038/nature08250
Contact: Jay Shendure, University of Washington, Seattle, Wash.
e-mail: shendure@u.washington.edu

COMPANIES AND INSTITUTIONS MENTIONED

Complete Genomics Inc., Mountain View, Calif.
Illumina Inc. (NASDAQ:ILMN), San Diego, Calif.
Knome Inc., Cambridge, Mass.
New York University, New York, N.Y.
University of Washington, Seattle, Wash.