Sequencing human diversity

By Lev Osherovich, Senior Writer

A team led by researchers at **GlaxoSmithKline plc** and the **University of California, Los Angeles** has amassed the most extensive catalog to date of sequence variation in genes that encode drug targets.¹ Unraveling how these variants influence drug response and disease susceptibility will require phenotypic studies in model systems.

Previously, the GSK team assembled a large database of single-nucleotide polymorphisms across the genomes of healthy individuals.^{2,3} That work painted a picture of human genetic diversity, but the microarray technology used in the study was not suited for identifying functional differences in disease-related genes.

To pinpoint those differences, the GSK team turned to a higher-resolution technique—DNA sequencing.

In the new study, the team sequenced the entirety of 202 drug target genes in a total of 14,002 cases and controls from a dozen cohorts of patients with cardiovascular, metabolic, autoimmune and neurological diseases.

The group expected to uncover a few key mutations present in patients with disease. Instead, they found many patients and controls harbored multiple genetic variants that differed from reference DNA sequences obtained from earlier sequencing efforts such as the Human Genome Project and the 1000 Genomes project.

“We asked what is the impact of rare genetic variants on disease in a large scale,” said team coleader Matthew Nelson, director of statistical genetics

at GSK. “What we had hoped to find was a few genes with genetic variants that are associated with disease. In fact, most patients had a large variety of rare mutations.”

“We found surprising heterogeneity across drug target genes,” added Stephanie Chissoe, acting head of genetics at GSK.

The study was co-led by John Novembre, assistant professor of ecology and evolutionary biology at UCLA.

“The sheer quantity and abundance of rare variants sprinkled

throughout the genome is a surprise,” said Joshua Akey, associate professor of genome sciences at the **University of Washington**.

Akey heads an academic consortium that has conducted a survey of genetic variation across the exome, the portion of the genome that encodes proteins.⁴

Whereas the GSK-UCLA team went deep into drug target genes in a large number of individuals, Akey’s team went wide and sequenced 15,585 genes from 2,440 individuals. Both teams reported their findings concurrently in *Science*.

Everybody is different

Nelson’s team sequenced 864 kilobase pairs and found that 1 in 17 nucleotides harbored at least 1 mutation compared with a reference sequence. The more subjects the team sequenced, the more variants they found. Extrapolating to a million individuals, the team predicted there would be up to 452 variants per kilobase pair of sequenced DNA.

Most of the variants found by the sequencing effort were rare—74% of the mutations were found in only 1–2 subjects. Moreover, 90% of the mutations had never been reported.

Some of the mutations were silent, but a fraction of them changed the amino acid sequence and presumably the function of the encoded protein. Indeed, 105 protein-altering variants in 73 genes were found in multiple individuals.

Akey said his own team’s analysis of the prevalence of rare genetic variants in the wider genome is in line with what the GSK and UCLA team saw. His team’s data predict the average individual’s exome contains 150 variants that deviate from the reference sequence.

Variation information

The findings suggest most individuals are likely to carry a handful of rare genetic variants that affect the function in disease-associated proteins. However, the complex nature of diseases affected by these genes makes it difficult to predict how—and if—the mutations affect disease or drug susceptibility.

“The biggest challenge in the field is to show a causal relationship between these rare variants and disease,” said Akey. “Sometimes these variants can obviously affect the structure or function of proteins, but most of the time these effects will be not obvious.”

Nelson said the scale of his group’s study was not sufficient to prove that a particular gene variant causes disease or affects drug response. This is because the majority of the mutations are unique or too rare to support a statistical argument for causality.

Instead, Nelson and Chissoe think the best use of the data will be to guide the design of experiments to test how these variants affect protein function or drug response in preclinical disease models.

“There’s reason to believe that the functional impact of these rare variants can shed light on the action of drugs that hit these targets,” said Chissoe. “One possibility is to test the effect of the rare variants in a model system. Another idea is to look at how a drug works” on proteins encoded by the mutant genes.

As an example of how variation affects protein function, Nelson cited one of the genes sequenced by his team, *lipoprotein-associated phospholipase A₂* (PLA₂G7; PAFAH; Lp-PLA₂). PLA₂G7 is the target of GSK and **Human Genome Sciences Inc.**’s darapladib, which is in Phase III

“The biggest challenge in the field is to show a causal relationship between these rare variants and disease. Sometimes these variants can obviously affect the structure or function of proteins, but most of the time these effects will be not obvious.”

—Joshua Akey,
University of Washington

testing for atherosclerosis and coronary artery disease (CAD).

In an earlier pilot study of sequence variation in *PLA₂G7*, Nelson's team identified 8 rare variants in a cohort of 2,000 European individuals. *In vitro* studies showed that these variants reduced the enzyme's activity.⁵ Because homozygous carriers of the mutations have low *PLA₂G7* activity, the team predicted that those rare individuals would likely not benefit from darapladib treatment and should be excluded from trials of the compound.

Nelson and Chisoe said studying the disease phenotypes of individuals who carry such rare variants also could help validate potential drug targets, as mutations that reduce the protein's function might predict the effect of pharmacologically inhibiting those proteins.

Akey said that as more sequencing data roll in, the relationships between gene variants and diseases will become more apparent. If done with a sufficiently large number of individuals, direct sequencing could supersede indirect disease gene hunting methods like genomewide association studies.

"The simple approach is to collect a large number of patients and controls and ask whether variants in a particular gene correlate with disease," said Akey. "There will be an avalanche of studies coming out in the next year on these relationships."

Nelson said the sequence data from the study are not patented and will be made publicly available.

Osherovich, L. *SciBX* 5(22); doi:10.1038/scibx.2012.567

Published online May 31, 2012

REFERENCES

1. Nelson, M.R. *et al. Science*; published online May 17, 2012; doi:10.1126/science.1217876
Contact: Matthew R. Nelson, GlaxoSmithKline plc, Research Triangle Park, N.C.
e-mail: matthew.r.nelson@gsk.com
Contact: John Novembre, University of California, Los Angeles, Calif.
e-mail: jnovembre@ucla.edu
2. Nelson, M.R. *et al. Am. J. Hum. Genet.* **83**, 347–358 (2008)
3. Osherovich, L. *SciBX* **1**(38); doi:10.1038/scibx.2008.917
4. Tennessen, J.A. *et al. Science*; published online May 17, 2012; doi:10.1126/science.1219240
Contact: Joshua M. Akey, University of Washington, Seattle, Wash.
e-mail: akeyj@uw.edu
Contact: Michael J. Bamshad, same affiliation as above
e-mail: mbamshad@u.washington.edu
5. Song, K. *et al. Pharmacogenomics J.*; published online May 24, 2011; doi:10.1038/tpj.2011.20

COMPANIES AND INSTITUTIONS MENTIONED

GlaxoSmithKline plc (LSE:GSK; NYSE:GSK), London, U.K.
Human Genome Sciences Inc. (NASDAQ:HGSI), Rockville, Md.
University of California, Los Angeles, Calif.
University of Washington, Seattle, Wash.