

MAY 10, 2007

Cataloging the Structural Variations in Human Genetics

A major new effort to uncover the medium- and large-scale genetic differences between humans may soon reveal DNA sequences that contribute to a wide range of diseases, according to a paper by Howard Hughes Medical Institute investigator Evan Eichler and 17 colleagues published in the May 10, 2007, *Nature*. The undertaking will help researchers identify structural variations in DNA sequences, which Eichler says amount to as much as five to ten percent of the human genome.

Past studies of human genetic differences usually have focused on the individual letters or bases of a DNA sequence. But the genetic differences between humans amount to more than simple spelling errors. Structural changes — insertions, duplications, deletions, and inversions of DNA — are extremely common in the human population, says Eichler. In fact, more bases are involved in structural changes in the genome than are involved in single-base-pair changes.

"It's a lot of work, because it's essentially doing 62 additional human genome projects."

— **Evan E. Eichler**

In some cases, individual genes appear in multiple copies because of duplications of DNA segments. In other cases, segments of DNA appear in some people but not others, which means that the reference human genome produced by the Human Genome Project is incomplete. We're finding new sequence in the human genome that is not in the reference sequence, Eichler says.

These structural changes can influence both disease susceptibility and the normal functioning and appearance of the body. Color-blindness, increased risk of prostate cancer, and susceptibility to some forms of cardiovascular disease result from deletions of particular genes or parts of genes. Extra copies of a gene known as *CC3LI* reduce a person's susceptibility to HIV infection and progression to AIDS. Lower than normal quantities of other genes can lead to intestinal or kidney diseases.

Variation in the number of genes or in gene regulation caused by structural rearrangements may also contribute to more common diseases. The million

dollar question is what is the genetic basis of diseases like diabetes, hypertension, and high cholesterol levels? says Eichler. We know there is a genetic factor, but what is the role of single base pair changes versus structural changes?

The project Eichler and his colleagues describe in their paper will help answer this question. Using DNA from 62 people who were studied as part of the International HapMap Project, they are creating bacterial libraries of DNA segments for each person. The ends of the segments are then sequenced to uncover evidence of structural variation. Whenever such evidence is found, the entire DNA segment is sequenced to catalog all of the genetic differences between the segment and the reference sequence.

The result, says Eichler, will be a tool that geneticists can use to associate structural variation with particular diseases. It might be that if I have an extra copy of gene A, my threshold for disease X may be higher or lower. Geneticists will then be able to test, or genotype, large numbers of individuals who have a particular disease to look for structural variants that they have in common. If a given variant is contributing to a disease, it will occur at a higher frequency in people with the disease.

Knowing about structural variation in the human genome will also allow geneticists to analyze single-base-pair changes more effectively, according to Aravinda Chakravarti, a geneticist at The Johns Hopkins University School of Medicine who was not a coauthor of the paper. We have to look at structural variants from a different perspective, because they are adding or subtracting something from the genome, Chakravarti says. By understanding the patterns of both structural variants and single-base-pair changes in the population, we'll learn a lot. To use both kinds of information in tandem, Eichler and his colleagues plan to incorporate the structural information they gather into existing databases on single-base-pair changes.

The project, which is being funded by the National Human Genome Research Institute at the National Institutes of Health, is difficult and expensive, Eichler admits. It's a lot of work, because it's essentially doing 62 additional human genome projects, he says. Having been involved in the first one, I swore I would never do it again. But in this case we're looking at the coolest parts of the genome.