



E-mail this page

Oct. 28, 2010 | Health and Medicine

New methods detect subtleties in human genomes' repetitive landscapes

By Leila Gray

News & Community Relations

[enlarge](#)

Scientists have invented methods to scout the human genome's repetitive landscapes, where DNA sequences are highly identical and heavily duplicated. These advances, as reported yesterday in *Science*, can identify subtle but important differences among people in the number and content of repeated DNA segments.

This, and related research on the completion of the first map of human variation from the 1000 Genomes Project reported in *Nature*, were heralded by news media around the world as a "new era of human genetics."

Copy number variations partly account for the normal diversity among people. Copy number variations might also be why some people, and not others, have certain disorders or disease susceptibilities, and might also determine how severely they are affected.

Until about a year ago, locating and counting the number of duplicated copies of DNA sequences was almost impossible. The more copies of a duplicated gene that are present, the harder they are to assess accurately.

"These difficulties resulted in a lack of understanding of the true extent of human copy number variation," said Dr. Evan E.

Eichler, UW professor of genome sciences and senior author of the *Science* paper, "The most dynamic and variable genes are frequently excluded from genome-wide studies." These hard-to-study genes are also among the most interesting because of their suspected contributions to human evolution, brain development, metabolism and disease immunity.

Researchers in Eichler's lab have developed several analytical and computational techniques to overcome obstacles in looking at multicopy genes. The lead authors of the study are Peter H. Sudmant and Jacob O. Kitzman, both graduate students in the UW Department of Genome Sciences.

Working with colleagues in the 1000 Genomes Project and at Agilent Technologies, the UW group used the new techniques to deeply probe and compare the genomes of 159 individuals. In assessing the entire genomes of these individuals, the researchers were able to accurately assay previously intractable duplicated genes and gene families.

The researchers demonstrated that the methods could estimate correctly the absolute number of copies of segments as small as 1,900 DNA base pairs, and could count numbers of copies ranging from 0 to 48. A human genome is made up of about 3 billion DNA base pair. Each pair consists of two bonded molecules called nucleotides, the basic structural unit of DNA.

"We identified 4.1 million singly unique nucleotide positions informative in distinguishing specific copies," the authors reported. The researchers took this information to genotype the number of copies and the content of genes that had been duplicated to or more different positions on the genome thereby became free to function on their own. These duplicated genes reveal changes that occurred during evolution.

The data allowed the researchers to identify duplicated genes specific to humans, in comparison to apes like gorilla, orangutans, and chimps. The researchers observed that these duplications occurred in genes associated with brain development. These include genes implicated in the growth and branching of brain cell connections, in abnormally large or small head size, in a particular dopamine (reward signal in the brain) receptor, in visual-spatial and social deficits, in reducing the severity of spinal muscular atrophy, and in intellectual disability and epilepsy.

Copy number variations occur in only about 7 percent to 9 percent of human genes, the researchers found. Most of our genes come standard: two copies. Even among copy number variable genes, the researchers learned that 80 percent of them vary between 0 and 5 copies.



Dr. Evan E. Eichler in his UW genome sciences lab, where he studies segmental duplications of DNA associated with human evolution, population diversity, and genetic disorders. His lab team developed new methods to count and explore the content of these duplications.

"Extreme gene variation," the researchers noted, "is limited to only a few gene families." In this study, they identified 56 of the most variable gene families. These ranged in median copy number from 5 to approximately 368.

"These genes were dramatically enriched for segmental duplication," the researchers noted. Segmental duplications are regions that were originally identified in the Human Genome Project as long, repeated blocks of the genome.

The researchers report discovering about 44 "hidden" members of duplicated gene families never before identified in the reference model of the human genome.

"The missing members of these gene families," the researchers suggested, "should be targeted for sequence finishing in order to more accurately capture the architecture and diversity of the human genome."

While duplications of segments of the genome appear to have led to many of the qualities that distinguish human beings from other primate species, areas of the genome in which duplications promote recurrent rearrangements have also been associated with debilitating diseases like intellectual disability, schizophrenia and autism.

The researchers hypothesize, "Extreme variation resulting from duplications may contribute to genomic instability associated with disease."

Overall, the results of the study shows scientists can now leverage newly developed techniques to explore some of the most complex genetic regions of the human genome. Still, a portion of the genome remains impenetrable. About 28 large regions of the human genome have such extraordinary complexity that as yet it is impossible to interpret the underlying pattern of genetic diversity, the authors said.

Despite this limitation, the approaches tested in the study hold promise for improving the understanding of how copy number variation contributes to human health and illness.

"Our approach," the researchers concluded, "makes many of the highly duplicated regions of the human genome -- and the more than 1,000 previously inaccessible human genes that lie therein -- accessible to genetic studies of disease association."

###

In addition to Eichler, Sudmant, and Kitzman, the scientists on the study are Francesca Antonacci, Can Alkan and Maika Malig, all from the UW Department of Genome Sciences; Anya Tsalenko, Nick Sampas, and Laurakay Bruhn, all of Agilent Technologies in Santa Clare, Calif., Jay Shendure, UW assistant professor of genome sciences, and participants and institutions in the 1000 Genome Project.

The work was supported by a Natural Sciences and Engineering Research Council of Canada Fellowship, a National Science Foundation Fellowship, and a grant from the National Human Genome Research Institute, National Institutes of Health. Eichler is an investigator of the Howard Hughes Medical Institute.

[Log in to UW News+Community to add a comment.](#)

