

MAY 01, 2008

New Map Reveals Dynamic Variation in Human Genome

A team of researchers led by Howard Hughes Medical Institute investigator Evan E. Eichler at the University of Washington has produced the first high-resolution map showing the structural variation that exists in the human genome. With the map, researchers can now begin to see how the underlying structure of one person's genome differs from that of another.

Eichler and a team of 45 colleagues examined the complete DNA sequences of eight people: four of African descent, two of Asian descent, and two of western European descent. They compared the DNA sequence of those eight people to the DNA sequence derived from the Human Genome Project, which is known as the reference sequence.

"Having eight new genomes like this will allow us to benchmark what we can and cannot detect."

— **Evan E. Eichler**

The resulting picture of the genome is much more complex than geneticists envisioned just a few years ago, but this complexity is vital. These are the details we need to find further associations with disease, Eichler said.

In a research article published on May 1, 2008, in the journal *Nature*, Eichler and his colleagues analyzed long variant stretches of DNA ranging from a few thousand to a few million base pairs in length. The new information uncovered in the studies will help researchers understand how humans are genetically different from one another. This is the first time that individual variants have been comprehensively cloned and sequenced to high quality, said Eichler. That information suggests mechanisms for genetic change that previously could not be inferred.

Geneticists traditionally have focused on changes in single letters — or base pairs — of a DNA sequence. But in recent years, several groups of researchers — including Eichler's group — have demonstrated that some of the most important genetic differences between humans involve larger segments of DNA.

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.

In various parts of our genome, some people have segments of DNA sequence that other people do not have. In other parts of our DNA, large genetic regions may be flipped in one person compared with another. These genetic differences can influence a person's susceptibility to heart disease, autism, lupus, HIV infection, and many other diseases.

Across all nine genomes analyzed in the *Nature* article, the researchers found 1,695 regions where people had DNA insertions, deletions, or inversions more than about 6,000 base pairs long. In some of these locations, all nine of the genomes were structurally different. At other sites, just one or a few people had structural variants.

The new analyses showed that the eight new genomes studied also have 525 segments of DNA that are not in the original reference genome. The size of those segments range from a few thousand to 130,000 base pairs. These results strongly argue that the human genome sequence is still incomplete, Eichler and his colleagues write in their paper. The authors suggest that it will be necessary to sequence additional genomes to fill the remaining gaps.

Taking a closer look, the researchers sequenced 261 of the structural variants letter by letter. In these regions, they found many structural variants smaller than 6,000 base pairs, ranging all the way down to insertions, deletions, or inversions of just a few base pairs. For unknown reasons, some parts of our genome are much more variable than others. For example, areas containing genes that are involved in the structural integrity of the body, such as the skin or the lining of the gut, are remarkably diverse. That was surprising to us, Eichler said.

An analysis of the 1,695 variable regions reveals that many occur where segments of DNA are repeated. These repeated segments of DNA have a tendency to misalign during the process that produces sperm and egg cells, which results in insertions and deletions of DNA. Roughly half of the insertions and deletions appear to be caused by that mechanism, said Jeffrey M. Kidd, a graduate student in Eichler's laboratory who was the lead author of the paper. To assign these events to misalignments, you need high-quality data.

Understanding these mutational processes will be critical as an increasing number of human genomes are sequenced. 2008 will be a big year for sequencing genomes, Eichler said. The eight people that Eichler's team studied are part of a much larger group whose genomes will be sequenced as part of the 1,000 Genomes Project, an international effort to sequence the genomes of people from around the world. Having eight new genomes like this will allow us to benchmark what we can and cannot detect.

Understanding structural variation also is essential in developing new technologies designed to detect the genetic differences among people. For example, so-called SNP chips, whether used in research or in clinical applications, need to reflect this structural variation to find links between particular gene variants and diseases. If you depended on the latest and greatest chip, you wouldn't find an association for about 50 percent of these sites, said Eichler.

Besides their potential applications, the new results provide a wealth of data to explore hypotheses and make discoveries, according to Eichler. What's exciting to me is that we now have, in essence, eight new reference human genomes.