OPINION

Mapping diversity

BY **APOORVA MANDAVILLI**

1 MAY 2008

dfac7a5c-8762-04e4-95e0-e79db1238414.jpg

When scientists initially talked about how genetically diverse people are, they focused on SNPs, which are sequence differences of a single base.

But it?s become increasingly clear that much of our diversity in fact derives from structural variations: large-scale differences in which stretches of sequence ranging from a few thousand to few million bases are lost, gained, or inverted.

Researchers yesterday released the first high-resolution map of structural variation across the human genome. The paper, published in ***Nature***, represents the first time that individual variants have been cloned and sequenced to high quality.

Among autism researchers, these variations are commonly referred to as copy number variations (CNVs) ? and their link to many diseases, including autism and schizophrenia, is **steadily gaining traction**.

As part of the **Human Genome Structural Variation Project**, **Evan Eichler** and 45 colleagues compared eight human genomes ?four of Yoruba Nigerian descent, two of Asian descent, and two of western European descent ? against the standard human reference genome completed in 2003. The eight samples had originally been collected as part of the **International HapMap Project**.

The researchers identified 1,695 CNV sites more than 6,000 base pairs long: 747 deletions, 724 insertions, and 224 inversions.

Roughly 40% of these CNVs had not been previously described; about half were found in at least two of the eight genomes analyzed.

Certain types of CNVs are more common on some chromosomes than on others. For example, the X chromosome is more prone to inversions than are other chromosomes. Some regions are unexpectedly diverse: for example, genes involved in the structural integrity of the body, such as the skin or the lining of the gut.

Interestingly, the same sites that are prone to CNVs also seem to show higher rates of nucleotide variation. For example, one region on chromosome eight has both an unusually high rate of nucleotide variation and one of the highest concentrations of CNVs.

The team also found 525 regions containing new sequences not found in the reference genome, suggesting that the reference genome itself might need updating.

The researchers plan to analyze up to 12 more individuals, whose genomes will be sequenced as part of the **1,000 Genomes Project**.