New map of genomic variations will enable disease research

July 15 2009

Genetics researchers have unveiled a powerful new resource for scientists and health providers studying human illnesses--a reference standard of deletions and duplications of DNA found in the human genome. Drawn from over 2,000 healthy persons, the study provides one of the deepest and broadest sets of copy number variations (CNVs) available to date, along with a new research tool for diagnosing and identifying genetic problems in patients.

A team from The Children's Hospital of Philadelphia published its high-resolution map and analysis of CNVs in the human genome in the July 10 online edition of the journal *Genome Research*.

In contrast to alterations to a single base of DNA, which are single nucleotide polymorphisms, or SNPs, often referred to as "snips," CNVs are larger variations in DNA structure. As changes to a single DNA letter, SNPs might be considered misspellings or alternate spellings of a word, while CNVs are losses of whole phrases, paragraphs or even pages (deletions), or are repeated sections (duplications). Some CNVs are inserted stretches of DNA from other parts of the genome. Both SNPs and CNVs contribute to genetic diversity and disease by changing the action of genes for which DNA carries coded instructions.

"We all carry a number of these variations in our own genomes," said study co-leader Peter S. White, Ph.D., a molecular geneticist and director of the Center for Biomedical Informatics at Children's Hospital. "Some CNVs contribute to a disorder, but most of them do not, and it is

often challenging to determine which are important. One approach is to compare CNVs in healthy individuals to those in patients with a disease, to find those CNVs that seem to occur primarily in people with a certain disease. Our map provides a large and uniform baseline standard to indicate which CNVs represent normal variation."

The investigators analyzed DNA from blood samples taken from 2,026 subjects. The subjects were healthy children and their parents, all of them drawn from primary care and well-child clinics in the Children's Hospital health care network. Of the samples, 65 percent were from Caucasians and 34 percent from African Americans. The number of subjects makes this CNV collection among the largest reported to date, and because all the samples were collected and analyzed under the same protocols, using the same technology, and at one institution, the results have a uniformity that increases their value as benchmarks.

The CNV map has a higher resolution than most previous efforts, say the authors, with over 50,000 CNVs catalogued throughout the genome. Three-quarters of these were "non-unique," occurring in multiple unrelated individuals. A majority (51.5 percent) of these non-unique CNVs were newly discovered. On average the healthy subjects in the study have approximately 27 CNVs each.

The researchers have posted the full CNV database on the Hospital's website, where it is freely available in searchable form to gene researchers worldwide. The web browser also enables researchers to compare specific CNVs to those collected in public data repositories from other institutions.

"This resource will be very important in enabling rapid and accurate diagnoses of rare diseases resulting from CNVs," said lead author Tamim H. Shaikh, Ph.D., a molecular geneticist at Children's Hospital. Often puzzling to physicians, such genetic diseases may be individually

rare, but collectively occur at frequencies that are comparable to the incidence of well-known disorders such as Down syndrome. "In order to pinpoint the one CNV that is the cause of a disease, it is critical to quickly eliminate those that are part of the spectrum of normal variation that exists in the human genome. That's what this CNV data and other similar resources allow us to do," Shaikh added.

As an example of the clinical usefulness of their database, the authors analyzed DNA from a child with multiple congenital problems, including developmental delay and brain malformations. They found 35 CNVs, of which 32 were previously detected in healthy controls. Two of the patient's three unique CNVs were relatively small in size, but the third CNV was a deletion in chromosome 17 that encompassed 51 genes, including several that are active in early prenatal development. Unlike most of the other CNVs, it did not occur in the child's parents, strongly supporting the conclusion that the chromosome deletion arose spontaneously in the patient and that it caused the child's disease.

To detect CNVs in the thousands of samples, the investigators used highly automated gene-analyzing technology at the Center for Applied Genomics at Children's Hospital, directed by Hakon Hakonarson, M.D., Ph.D., a co-leader of this study. "Although these CNVs were detected in healthy children, they may have significant disease implications that may not manifest until later in life," said Hakonarson. Earlier this year, Hakonarson and colleagues published groundbreaking studies of CNVs in autistic spectrum disorders and attention-deficit hyperactivity disorder. Both studies found CNVs in gene regions involved in neurological development during early childhood.

The new database has another strength, added Shaikh. Because it analyzed large numbers of samples from both Caucasians and African Americans, it measured CNV levels that differ between the two ethnic groups, and enables clinicians to make more precise diagnoses. Shaikh

added that the researchers expect to expand the database with larger sample sizes and data from additional ethnic populations.

In addition to its use in diagnosis, said White, the database may also assist researchers studying molecular evolution, for example, those investigating how genetic variations occurred as human populations spread across continents.

Source: Children's Hospital of Philadelphia (news : web)