Rare genetic variants create 'synthetic' genome-wide signals of disease risk

January 26 2010

Scientists at Duke University Medical Center say they are now convinced that rare genetic variants - as opposed to more common ones - lie at the heart of the genetic component of most common diseases.

The finding appearing in *PLoS Biology,* stems from a series of simulation studies, and challenges common interpretations of a multitude of genome-wide association studies (GWAS) that have identified thousands of single-DNA letter changes associated with greater risk of common diseases such as cancer, heart disease and diabetes. It may also hold important implications for understanding the underlying architecture of human disease.

Over the past several years, the research community has used tests sensitive enough to detect single-letter differences in DNA (called SNPs) to zero in on common human gene variants and their relationship to disease. Although thousands of significant associations have been identified - some of which led *Science* Magazine to declare human genetic variation "the discovery of the year" in 2007 - the precise genetic changes responsible for these observations have rarely been found.

"It may be because we have been looking in the wrong place," says David Goldstein Ph.D., director of the Center for Human Genome Variation in the Duke Institute for Genome Sciences & Policy (IGSP).Goldstein has long argued that rare, more powerful genetic variation lies in more remote stretches of the genome that GWA studies don't cover.

In GWA studies, researchers compare what happens to people who carry the alterations to those who do not carry them. The SNPs themselves aren't necessarily seen as the causative agents of disease but instead point to a region in the genome - generally one close by - where genetic alteration causing the disease might reside. GWA studies are driven by the notion that finding common variation in specific diseases would point to the specific genes at fault and eventually to corrective targeted therapies and a new era of personalized medicine.

"The trouble is, that just hasn't happened," says Goldstein. "Of course, it is still early and there have been successes. But for most common diseases, the common variants implicated account for only a small proportion of the genetic component. We are now pretty sure that much of the so called 'missing heritability' lies within the huge class of relatively or very rare genetic variants which were not represented in previous studies."

Goldstein, along with lead author Samuel Dickson, Ph.D., a bioinformatician in the IGSP, simulated case-control studies with sample sizes between 2,000 and 6,000 subjects by using simulated genealogical trees to create a realistic spread of variants across one section of a chromosome. They then used statistics and computer modeling to assess the possibility that a common variant, like those used in GWA studies, would be associated with a disease caused by one or more rare variants in the region.

They found that about a third of the simulations revealed an association with a common variant. They also found that these "synthetic" associations grew stronger as the number of rare variants increased and the relationships remained stable even when computer modeling allowed for chromosomal recombination - mimicking what happens in the real world over time. "Basically we showed that not only is it possible that rare variants are behind many of the results of recent findings, but that

there are likely to be many more to be found as researchers shift their focus to methods that will find rare variants," says Dickson.

Researchers say their findings are supported by real data from patients with sickle cell disease and a genetically linked form of hearing loss, two disorders representing the possible extremes of synthetic association. Sickle cell anemia is caused by mutations in one gene. It occurs in one in every 600 African Americans in the U.S. In comparison, hearing loss is a complex disorder affecting one in every 1000 newborns in the U.S. that involves more than two dozen genes and hundreds of harmful mutations in those genes.

The scientists ran GWA studies among a small number of cases and thousands of controls for both diseases. They found that common gene variants in the region showed genome wide significance, even though the only causal sites were known to be rare variants with a much bigger impact on disease, as predicted by the computer simulations.

"This tells us that we will surely need to turn to more comprehensive whole genome sequencing studies of more carefully selected subjects if we want to discover more meaningful relationships between genetic variation and disease," says Goldstein. "While such studies are undoubtedly more complex, expensive and time-consuming, we really have no choice if we want to deepen our knowledge about the genetic underpinnings of human disease."

Provided by Duke University Medical Center