Biostatisticians identify genes linked to heart disease

February 6 2013

Recently, large studies have identified some of the genetic basis for important common diseases such as heart disease and diabetes, but most of the genetic contribution to them remains undiscovered. Now researchers at the University of Massachusetts Amherst led by biostatistician Andrea Foulkes have applied sophisticated statistical tools to existing large databases to reveal substantial new information about genes that cause such conditions as high cholesterol linked to heart disease.

Foulkes says, "This new approach to data analysis provides opportunities for developing new treatments. It also advances approaches to identifying people at greatest risk for heart disease. Another important point is that our method is straightforward to use with freely available computer software and can be applied broadly to advance genetic knowledge of many diseases. We hope this moves us toward greater understanding of common disorders and improving overall health in our society."

The new analytical approach she developed with cardiologist Dr. Muredach Reilly at the University of Pennsylvania and others is called "Mixed modeling of Meta-Analysis P-values" or MixMAP. Because it makes use of existing public databases, the powerful new method represents a low-cost tool for investigators.

Foulkes, who directs the Institute for Computational Biology, Biostatistics and Bioinformatics at UMass Amherst, explains that

MixMAP draws on a principled statistical modeling framework and the vast array of summary data now available from genetic association studies to formally test at a new, locus-level, association. Other members of the team at UMass Amherst are Rongheng Lin, an assistant professor, and postdoctoral researchers Gregory Matthews and Ujwall Das.

The new method, which is generalizable to other problems, takes into account the structure of the genome, Foulkes explains. In genetic epidemiology, traditional genome-wide association studies look at common genetic variations in a group of people with a disease compared to healthy controls, focusing on single-nucleotide polymorphisms (SNP) to see if any single variant is associated with that disease. If one variant is found more often in people with the disease, the SNP is thought to be associated with it.

While that traditional statistical method looks for one unusual "needle in a haystack" as a possible disease signal, Foulkes and colleagues' new method uses knowledge of DNA regions in the genome that are likely to contain several genetic signals for disease variation clumped together in one region. Thus, it is able to detect groups of unusual variants rather than just single SNPs, offering a way to "call out" gene regions that have a consistent signal above normal variation.

Foulkes offers a non-technical analogy, "It's like listening to an orchestra. If there is one drum, we all hear it, and we hear the cello section even if each instrument is playing quietly, because they are all playing together. Current statistical methods can hear the drum, but not the cellos, no matter how many are playing, because no single cello is as loud as the drum. Our method can do this, which adds considerable statistical strength to our ability to identify new genes that may be associated with disease."

Reilly points out that as a complementary strategy to traditional methods,

MixMAP will be of great interest to the statistical and cardiovascular genomics community. It can be applied to existing genetic studies of blood cholesterol levels and it already suggests a dozen new genes to explore further for an important heart disease risk factor: Low density lipoprotein cholesterol.

Foulkes characterizes the new technique as discovery science still in need of validation, but it's discovery science that goes farther than usual by using sophisticated modeling approaches to quantify error. "We've done better than simply identify the strongest signals, we've quantified measures of association to show they are statistically meaningful," she points out.

Overall, the authors say, "MixMAP offers new and complementary information as compared to single nucleotide polymorphism-based analysis approaches and is straightforward to implement with existing open-source statistical software tools."

The research is published today in *Plos One*.

Provided by University of Massachusetts Amherst