Researchers create powerful new method to analyze genetic data

June 12 2012

University of Texas Medical Branch at Galveston researchers have developed a powerful visual analytical approach to explore genetic data, enabling scientists to identify novel patterns of information that could be crucial to human health.

The method, which combines three different "bipartite visual representations" of genetic information, is described in an article to appear in the Journal of the American Medical Informatics Association. The work won a distinguished paper award when it was presented at the AMIA Summit on Translational Bioinformatics in March 2012.

In the paper, the authors use their technique to analyze data on genetic alterations in humans known as single-nucleotide polymorphisms, or SNPs. Among other things, the frequencies of particular SNPs are associated with an individual's ancestral origins; for the study, the researchers chose to examine SNP data from 60 individuals from Nigeria and 60 individuals from Utah.

"We selected SNPs that we already knew differentiated between the two groups, and then showed that our method can reveal more about the data than traditional methods," said UTMB associate professor Suresh Bhavnani, lead author on the JAMIA paper and a member of UTMB's Institute for Translational Sciences. "This is a fresh way of looking at genetic data, a methodological contribution that we believe can help biologists and clinicians make better sense of a variety of biomarkers."

Like many kinds of biomedical data, Bhavnani said, datasets describing individuals and their SNPs are particularly suited to visual representations that are bipartite: that is, they simultaneously present two different classes of data. In the case of the Utah-Nigeria SNP data, Bhavnani and his colleagues started with what is known as a bipartite network visualization — an intricate computer-generated arrangement of colored dots and black, gray and white lines.

"In the bipartite network you see both the individuals and their genetic profiles simultaneously, and cognitively that's really important," Bhavnani said. "You can look at the individuals and know immediately which SNPs make them different from others, and conversely you can look at the SNPs to see how they are co-occurring, and with which individuals they are co-occurring. This rich representation enables you to quickly comprehend the complex bipartite relationships in the data"

The bipartite network visualization of the Utah-Nigeria individual-SNP data has distinct clusters on its left and right sides that correspond to the Utah and Nigerian subjects and SNPs. It also accurately portrays a genetic phenomenon called admixture, in which an individual possesses SNPs that are characteristic of individuals from Utah as well as from Nigeria. Admixed individuals are placed on the edges of their clusters, relatively close to the center of the visualization. The identification of admixed individuals, and the implicated SNPs could help in the design of case-control studies where there is a need for the selection of homogenous sets of individual from different ancestral origins.

To produce an even more detailed picture of the individual-SNP information, the researchers applied two other bipartite visualization techniques to the data: the bipartite heat map, and the bipartite Circos ideogram. In the heat map, rectangular cells laid out in a spreadsheet-like arrangement and colored white, gray, or black helped precisely define the boundaries of the clusters by clarifying individual-SNP relationships.

In the Circos ideogram, individuals and SNPs placed around the perimeter of a circle and linked with curved lines, enabling the researchers to more closely examine the admixed individuals' ties to SNPs in the clusters associated with both Utah and Nigeria.

"The network representation is very powerful because it gives you the overall structure of the data, but to really understand the complex relationships, you need these additional bipartite representations," Bhavnani said.

The JAMIA paper, according to Bhavnani, represents a proof of concept for the researchers' novel combination of methods, which can be applied to a wide range of biomedical questions. "You can think of anything – for example you could examine cases and controls in Alzheimer's disease, or you could compare children who are prone to ear infections and those aren't prone," Bhavnani said. "Whatever your disease or trait of interest is, our approach can handle it."

Provided by University of Texas Medical Branch at Galveston