# New approach for efficient analysis of emerging genetic data

September 6 2012

(Medical Xpress)—With the ability to sequence human genes comes an onslaught of raw material about the genetic characteristics that distinguish us, and wading through these reserves of data poses a major challenge for life scientists. Researchers at Roswell Park Cancer Institute (RPCI) and the Center for Human Genome Variation at Duke University Medical Center (DUMC) have developed an approach for analyzing data that can help researchers studying genetic factors in disease to quickly cull out relevant genetic patterns and identify variants that lead to particular disorders.

The researchers outline this new approach in a study published in the September issue of *The* American Journal of Human Genetics. They note that while genome-wide association studies (GWASs), which analyze the DNA of many people in order to reveal genetic variations associated with disease, have reported thousands of variants associated with different traits, it's very difficult to isolate "causal variants," those genetic irregularities that appear to indicate the presence of a particular disease.

RPCI biostatician Qianqian Zhu, PhD, Assistant Member of the Department of Biostatistics & Bioinformatics and Director of the Institute's Statistical Genetics & Genomics Resource, began this work as a postdoctoral researcher at Duke and completed the study after joining the Roswell Park faculty. The computational method that she and her collaborators established—which they call the preferential linkage disequilibrium approach—follows variants reported by GWASs, then

cross-references those variants with a comprehensive variant catalog generated through robust "next generation" sequencing in order to pinpoint causal variants. For this study, the team looked at DNA from 479 people of European descent.

"To test our method, we ran it on five diseases for which the causal variants are known, and in every case we did identify the real causal variant," notes Zhu, the paper's first author. "We're confident that our method can be applied to genome-wide association studies related to diseases for which there are no known causal variants, and by extension may advance the development of targeted approaches to treating those diseases."

"This approach helps to intergrade the large body of data available in GWASs with the rapidly accumulating sequence data," adds David B. Goldstein, Richard and Pat Johnson Distinguished University Professor and Director of the Center for Human Genome Variation at DUMC and senior author of the paper.

The study, "Prioritizing Genetic Variants for Causality on the Basis of Preferential Linkage Disequilibrium," is available online at www.cell.com/AJHG/abstract/S0002-9297%2812%2900367-9 .

Provided by Roswell Park Cancer Institute