

Big data allows computer engineers to find genetic clues in humans

March 27 2015



Big data: It's a term we read and hear about often, but is hard to grasp. Computer scientists at Washington University in St. Louis' School of Engineering & Applied Science tackled some big data about an important protein and discovered its connection in human history as well as clues about its role in complex neurological diseases.

Through a novel method of analyzing these big data, Sharlee Climer, PhD, research assistant professor in computer science, and Weixiong Zhang, PhD, professor of computer science and of genetics at the School of Medicine, discovered a region encompassing the gephyrin gene on chromosome 14 that underwent rapid evolution after splitting in two completely opposite directions thousands of years ago. Those opposite directions, known as yin and yang, are still strongly evident across

different populations of people around the world today.

The results of their research, done with Alan Templeton, PhD, the Charles Rebstock professor emeritus in the Department of Biology in the College of Arts & Sciences, appear in the March 27 issue of *Nature Communications*.

The gephyrin protein is a master regulator of receptors in the brain that transmit messages. Malfunction of the protein has been associated with epilepsy, Alzheimer's disease, schizophrenia and other [neurological diseases](#). Additionally, without gephyrin, our bodies are unable to synthesize an essential trace nutrient.

The research team used big data from the International HapMap Project, a public resource of genetic data from populations worldwide designed help researchers find genes associated with human disease, as well as from the 1000 Genomes project, another public data source of sequenced human genomes. In total, they looked at the genetic data from 3,438 individuals.

When they analyzed the data, they made an interesting discovery in a sequence of markers, called a haplotype, enveloping the gephyrin gene: up to 80 percent of the haplotypes were perfect yin and yang types, or complete opposites of the other. They were able to trace the split back to what is known as the Ancestral haplotype, or that of the most recent common human ancestor.

"We observed that the Ancestral haplotype split into two distinct haplotypes and subsequently underwent rapid evolution, as each haplotype possesses about 140 markers that are different from the Ancestral haplotype," Climer says. "These numerous mutations should have produced a large number of intermediate haplotypes, but the intermediates have almost entirely disappeared, and the divergent yin

and yang haplotypes are prevalent in populations representing every major human ancestry."

Using the data from the HapMap Project, they looked at the gephyrin region in several populations of people, including European, East and South Asian, and African heritage, and found variations in the haplotype frequencies of each of these populations. Those from African origin generally have more yang haplotypes, while those of European origin have more yin haplotypes. Those of Asian descent have nearly equal numbers of yin and yang haplotypes.

Humans carry pairs of chromosomes, and 30 percent of Japanese individuals carry two yin haplotypes or two yang haplotypes. Another 30 percent of these individuals possess both a yin and a yang haplotype, reflecting the roughly equal probability of inheriting either one.

To find this pattern within the huge datasets, the research team used a novel method to assess correlations between genetic markers called single nucleotide polymorphisms, or SNPs, which are variations in a DNA sequence that make humans different from each other.

The team's method, called BlocBuster, computes correlations between each pair of SNPs, then builds a network of those correlations. By observing the network, researchers can see clusters of correlated markers.

"For example, you could build a Facebook network using all of your Facebook friends," Climer says. "If two of your friends are friends with each other, you would connect them in the network. If you see that a cluster of people is interconnected with each other, they probably share something in common, such as a family relationship, a school, or some type of social interaction. Similarly, with an efficient algorithm and an adequate number of processors and time, we can look at every pair of

SNPs, build these networks and observe clusters of interconnected SNPs.

"The BlocBuster approach is a paradigm shift from the conventional methods for genome-wide association studies, or popularly known as GWAS, where one or a few markers were examined at a time," Zhang says. "It is truly a data mining technique for [big data](#) like those from HapMap and 1000 Genomes projects."

The researchers also can design this approach to look at complex traits and diseases.

"BlocBuster is able to detect combinations of networked genetic markers that are characteristic of complex traits," Zhang says. "It is suitable for analyzing traits, such as body weights, which are determined by multiple genetic factors, and genetic patterns in populations, such as the yin-yang haplotypes we discovered."

Ultimately, they expect this method will shed light on the genetic roots of disease.

"Most complex diseases arise due to a group of genetic variations interacting together," Climer says. "Different groups of people who get a disease may be affected by different groups of variations. There's not enough power to see most of these intricate associations when looking at single markers one at a time. We're taking a combinatorial approach—looking at combinations of markers together—and we're able to see the patterns."

More information: Climer S, Templeton A, Zhang W. Human gephyrin is encompassed within giant functional noncoding yin-yang sequences. *Nature Communications*, March 27, 2015. 6:6534 [DOI: 10.1038/ncomms7534](#) (2015).

Provided by Washington University in St. Louis

Citation: Big data allows computer engineers to find genetic clues in humans (2015, March 27)
retrieved 6 May 2024 from <https://phys.org/news/2015-03-big-genetic-clues-humans.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.