New 'genetic bar code' technique establishes ability to derive DNA information from RNA

April 8 2012

Researchers from Mount Sinai School of Medicine have developed a method to derive enough DNA information from non-DNA sources—such as RNA—to clearly identify individuals whose biological data are stored in massive research repositories. The approach may raise questions regarding the ability to protect individual identity when high-dimensional data are collected for research purposes. A paper introducing the technique appears in the April 8 online edition of *Nature Genetics*.

DNA contains the genetic instructions used in the development and functioning of every living cell. RNA acts as a messenger that relays genetic information in the cell so that the great majority of processes needed for tissue to function properly can be carried out.

To date, access to data bases with DNA information has been restricted and protected as it has long been considered the sole genetic fingerprint for every individual. However, vast amounts of RNA data have been made publicly available via a number of databases in the United States and Europe. These databases contain thousands of genomic studies from around the world.

In this study, lead authors Eric E. Schadt, PhD, and Ke Hao, PhD, developed a technique whereby a person's DNA could be inferred from RNA data using gene-expression levels monitored in any of a number of tissues. In contrast, most studies involving DNA and RNA begin with DNA sequences and then seek to associate expression patterns with

changes in DNA between individuals in a population. This is the first time going from RNA levels to DNA sequence has been described.

"By observing RNA levels in a given tissue, we can infer a genotypic barcode that uniquely tags an individual in ways that enables matching the individual to an independently derived DNA sample," said Dr. Schadt, Director of the Institute for Genomics and Multiscale Biology, the Jean C. and James W. Crystal Professor of Genomics, and Chair of the Department of Genetics and Genomics Sciences, Mount Sinai School of Medicine. "The potential uses for this information are significant. Not only can genotypic barcodes be deduced from RNA, but RNA levels in some tissue can inform not only individual characteristics like age and sex, but on diseases such as Alzheimer's and cancer, as well as the risks of developing those diseases."

Schadt adds, "The significance of our findings goes beyond medicine. For example, barcodes derived from individuals who participated in a research study, where RNA levels were monitored and deposited into publicly available databases, could be tested against DNA samples left at a crime scene as a way of identifying persons of interest."

Deducing a person's DNA sequence from gene expression patterns could have repercussions in health care and privacy. While specific laws and government regulations have been written to protect DNA-based information from misuse, it is unclear whether such laws apply to RNA—even though this study shows that RNA is informative at a deeper level compared to DNA regarding the current state of health of an individual.

"Rather than developing ways to further protect an individual's privacy given the ability to collect mountains of information on him or her, we would be better served by a society that accepts the fact that new types of high-dimensional data reflect deeply on who we are," Dr. Schadt said.

"We need to accept the reality that it is difficult—if not impossible—to shield personal information from others. It is akin to trying to protect privacy regarding appearances, for example, in a public place."

Dr. Schadt said he hopes the research will catalyze a discussion that might ultimately help resolve privacy debates, and encourage patients to provide data that will help their doctors better diagnose and treat their conditions. Increased access to, and greater quantities of, DNA and other biological information would also contribute to the greater good of medical science.

In the Nature Genetics study, Drs. Schadt and Hao, Associate Professor of Genetics at Mount Sinai School of Medicine, together with Sangsoon Woo, PhD, from the Department of Biostatistics at the University of Washington, analyzed RNA and DNA from 378 livers donated by European-Americans for transplant, as well as liver and adipose tissues from 580 people from the same population group undergoing gastric bypass surgery. The authors found that levels of RNA across many genes correlate with age, sex, body weight, and other risk factors for diseases like diabetes and heart disease, but then they also correlate in many cases with changes in DNA that are unique to a given individual.

The investigators used an algorithm that matches patterns of gene expression to variations at 1,000 single-DNA-base sites in the genome. It is an application of integrative biology that examines multiple dimensions of data (DNA and RNA) to better inform a given dimension (RNA).

"The relationship of DNA to RNA is like that of an orchestra and the symphony it plays," said Schadt describing the new technique. "The DNA (orchestra) remains the same, while the RNA pattern (quality of the music) changes in response to outside factors. The new technique is like hearing a symphony and deducing which instruments are in the

orchestra, essentially unwinding the developmental process to trace tissue samples back to RNA and the gene that instructed it."

Provided by The Mount Sinai Hospital