Researchers develop 'hi-def' copy number variation decoder

November 1 2010, By Paul Cantin

University of Toronto researchers have developed a new "high definition" computer program to analyze human DNA and more accurately detect genetic variants that affect individual traits like disease susceptibility and varying drug responses.

A multidisciplinary team combining computer science and biomedical data developed new theory and code to enable the precise determination of the number of copies of genes in the human genome. In the past few years, such copy number variation (or CNV) of genes has been shown to be a universal form of genetic variation and also found to cause diseases like autism and cancer, but until now it has remained very difficult to identify. This new computer algorithm, which promises to simplify CNV discovery, is unveiled in the cover story of this month's *Genome Research*.

"Most, but not all, genes occur in two copies in our genome, one inherited from each parent. When examining genome sequences we need to distinguish genes that may be present in zero or one copy or present in three or more copies and the complexity of DNA itself can make that very difficult," said Professor Michael Brudno of computer science and the Donnelly Centre for Cellular and Biomolecular Research, senior author.

Brudno, Canada Research Chair in Computational Biology, likens his new invention, called CNVer, to a game of "spot the difference," in this case searching for glitches in sub-microscopic pieces of DNA.

"Imagine two near-identical images - one photograph contains two cars, the other only one. If you cut those images into snippets and shuffled them (precisely what happens when you sequence DNA), it would be difficult to detect which image fragment belonged with the original picture. We have developed sophisticated methods to scrutinize connecting fragments around, or between, the vehicles, allowing both the number of cars (or copies of a gene) within the photograph and their location to be accurately reconstructed. Together this information allows us to see a high-definition view of the genome, while only looking at the individual small pieces," said Brudno.

Professor Stephen Scherer, director of the McLaughlin Centre U of T and the Centre for Applied Genomics at The Hospital for Sick Children who co-discovered genome-wide CNVs in 2004, commented: "This new tool will have tremendous impact in our ability to understand the medical relevance of CNVs in the massive amounts of data coming from personal genome sequencing projects."

Co-authors on this research are Paul Medvedev, Marc Fiume, Misko Dzamba and Tim Smith of the Department of Computer Science. Support for this research came from the Canadian Institutes of Health Research.

Provided by University of Toronto