October 27, 2009

# BME Researchers Help Advance Novel Personal Genotyping Technique

By Mark Dwortzan

To determine your personal predisposition to selected diseases, you can send a sample of your own DNA, along with a few hundred dollars, to one of several personal genotyping firms. Using a commercially available gene chip, a technician will then compare specific nucleotides in your DNA sample against hundreds of thousands of catalogued positions in the human genome, and pinpoint locations where the sample departs from the norm — and may indicate disease susceptibility.

*Michael Molla, a Center for Biodynamics fellow and post-doc in James Collins' lab, designed a novel microchip that detects variations in short, repeating DNA sequences throughout the human genome that could indicate predisposition to disease.*

Now a novel approach to genotyping that could provide additional personal genetic information has emerged from a team of scientists that includes three researchers in the Biomedical Engineering Department — Michael Molla, a Center for Biodynamics fellow and post-doc in Prof. James Collins' lab; Simon Kasif, professor of bioinformatics and biomedical engineering and director of the Computational Genomics Lab; and Prof. Charles Cantor of the Center for Advanced Biotechnology. The team reported on its innovation in the October 6 edition of Proceedings of the National Academy of Sciences.

## A New Approach to Genotyping

The new genotyping strategy exploits variations in the lengths of short DNA sequences called short tandem repeats (STRs) that appear over and over again in thousands of regions of the human genome. For example, in the short sequence of nucleotide bases "AGCAGCAGCAGCAGC," the triplet "AGC" occurs five times.

Individuals occasionally differ from the norm in their number of triplet copies within the short sequence — and thus the length of the sequence itself. Scientists have long linked abnormal length variations in STRs to Huntington's disease, schizophrenia and other major diseases, but until now have never investigated repeats at the whole genome level.

In the PNAS study, which compared STRs in three well-known versions of the human genome as well as that of a chimpanzee, Molla and his co-investigators discovered a high rate of length variation in STRs that extend beyond 20 nucleotides. Their finding suggests a new set of sequences in personal DNA samples to examine for disease susceptibility.

In the process, the researchers also introduced a new, inexpensive, whole-genome method to detect abnormal repeats that could be used to screen individuals' susceptibility to Huntington's and other diseases.

"Our whole genome repeat study is the first of its kind," said Molla, who specializes in computational methods for high-throughput biological data. "Our observations and proposed assay constitute a step toward understanding how these repeats work on a genome-wide scale and their implications for human health."

## A New Kind of Genetic Screening Chip

Applying computational analyses to existing sequence data, the research team compiled and assembled the first complete list of STRs in regions of human and chimp genomes where the repeated element is a triplet. For each of these regions, they compared the lengths of triplet repeats in the gene transcripts of the three human and one chimp genomes, and eventually observed a high rate of length variation in longer STRs.

The team also made a surprising discovery that could yield fresh insights about the nature of human DNA: Human genomes appear to have a strong preference for triplet repeats of length 11, 14, 17, or 20. "Dividing the lengths by three, we observed a remainder of two — one less than the

length of the repeated sequence — about 100 times more than a remainder of one or zero," noted Molla.

To find triplet STRs throughout human and chimp genomes more quickly and inexpensively in the future, Molla and his co-investigators proposed a DNA "capture array" with customized probes to adhere to all triplet repeats and their surrounding sequences in human and chimp genomes. Molla designed a new chip to perform the complex task.

Originally conceived by Kasif and Cantor, a pioneer in genomics technology, the PNAS study also includes major contributions from Kasif's first PhD student, Arthur Delcher, a senior research scientist at the Center for Bioinformatics and Computational Biology at the University of Maryland; and Shamil R. Sunyaev, Assistant Professor of Medicine and Health Sciences and Technology at Harvard Medical School and Brigham & Women's Hospital in Boston. This study was funded by the National Human Genome Research Institute, National Science Foundation and National Institutes of Health Informatics to Bedside (I2B2) Consortium.

The BU researchers and their co-investigators next plan to analyze more genomes, refine their genome screening assay and further verify their findings.

"We are hoping to scale up the assay technology to enable us to examine STR length variations in human individuals with a particular disease, such as cancer," said Kasif, "and to compare them to the normal population and check whether the individual has a predisposition to the disease."