# STRait RAZOR: A BIOINFORMATIC TOOL FOR LENGTH-BASED STR ALLELE-CALLING IN MASSIVELY PARALLEL SEQUENCING DATA

David H. Warshauer[1], David Lin[2], Kumar Hari[2], Ravi Jain[2], Carey Davis[1], Bobby LaRue[1], Jonathan L. King[1], and Bruce Budowle[1,3]
[1]Institute of Applied Genetics, Department of Forensic and Investigative Genetics, University of North Texas Health Science Center
[2]cBio, Inc., 37869 Abraham Street
[3]Center of Excellence in Genomic Medicine (CEGMR), King Abdulaziz University

Short tandem repeat (STR) interpretation makes up a substantial portion of the genetic analysis involved in forensic investigations. The current method of detection for these markers is size separation by capillary electrophoresis (CE). However, recent studies have shown that the rapidly-developing massively parallel sequencing (MPS) technologies also are capable of detecting STRs. MPS offers advantages over CE-based detection methods, such as the ability to type a substantially larger number of markers in a single run and to reveal intra-repeat and flanking region variations. While current MPS instrumentation can produce the necessary raw data, there are relatively few bioinformatic software packages available that can accurately interpret these data and produce standard STR allele calls.

This presentation will describe STRait Razor (the STR Allele Identification Tool – Razor), a newly developed Perl-based software tool that runs on the Linux/Unix operating system, which is designed to detect forensically-relevant STR alleles from FASTQ sequence data, based on allelic length. The results of analyses with STRait Razor are comparable with those of CE-based methods, while retaining the added informative value that MPS provides. STRait Razor is capable of analyzing STR loci with repeat motifs ranging from simple to complex without the need of virtual allelic ladders that require extensive allelic sequence data. It has been designed to interpret both single-end and paired-end data and relies on intelligent parallel processing to reduce analysis time. Users are presented with a number of customization options, including variable mismatch detection parameters and the ability to easily allow for the detection of alleles at new loci. In its current state, the software package detects alleles for 44 autosomal and Y-chromosome STR loci.

The efficiency and accuracy of STRait Razor were determined via concordance tests, using 5 whole blood samples. The samples were prepared for sequencing using two different library preparation chemistries, and were sequenced on two different Illumina® instruments. The allele calls made by STRait Razor were compared with the alleles determined by traditional CE methods. In total, 427 alleles were compared during this trial, and the results demonstrated that STRait Razor is capable of detecting STR alleles in raw sequencing data with 100% concordance. This presentation will summarize the results of this study, as well as emphasize concepts related to the effect of different preparation chemistries and sequencing parameters on the bioinformatic detection of STR alleles.