

# **MATCH/IT: A CONTINUOUS DNA INTERPRETATION METHOD THAT GENERATES LIKELIHOOD RATIO DISTRIBUTIONS**

Harish Swaminathan<sup>1</sup>, Catherine M. Grgicak<sup>2</sup>, Muriel Medard<sup>3</sup>, and Desmond S. Lun<sup>1</sup>

<sup>1</sup>Center for Computational and Integrative Biology, Rutgers University, Camden, NJ 08102, USA

<sup>2</sup>Biomedical Forensic Sciences Program, Boston University School of Medicine, Boston, MA 02118, USA

<sup>3</sup>Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

In recent years, continuous DNA interpretation methods have become available. These methods account for peak heights and various artifacts such as drop-out and stutter in a DNA profile when evaluating a match. As such, they are superior to binary methods (where peaks below the analytical threshold are removed, and the remaining are assumed to be true alleles or stutter) and semi-continuous methods (where an analytical threshold is used but drop-in and drop-out of alleles is also accounted for).

Despite the obvious benefits of continuous DNA interpretation, however, it also introduces new issues. Chief among them is the issue of selecting a mathematical model to describe the probability of observing a certain DNA profile given a sample with a certain genotype. While for binary methods the mathematical model is established and well-accepted, this is not so for semi-continuous and continuous methods. Indeed, various different models—all based on reasonable modeling assumptions—have been described in the literature for continuous DNA interpretation and are used by the methods available to date.

Statistics, such as the likelihood ratio (LR), calculated by continuous methods change as the mathematical model is changed, and it is therefore natural to ask, how meaningful is the LR obtained under a particular model? It has been proposed that this issue can be addressed by studying the LR distribution under alternative contributors, but efficiently computing the LR distribution for a continuous method is very challenging.

Here, we present *Match/It*, a continuous DNA interpretation method that efficiently computes both the LR for a given person of interest (POI) as well as the distribution of the LR under alternative, random POIs, and we discuss the results of testing *Match/It* on 307 1-, 2-, and 3-person experimental samples amplified using DNA target masses of 0.008 to 0.5 ng of DNA and the 29-cycle Identifiler™ Plus Amplification Kit protocol. We show that LRs do indeed vary significantly for the same POI on the same sample under different reasonable models and discuss how statistics such as the  $p$ -value and the Tippett test statistic  $P(\text{LR} > 1 | H_d)$  can be derived from the LR distribution to study its robustness.

A prototype of *Match/It* has been developed and will be released for research purposes.