Production of High-Fidelity Electropherograms Results in Improved and Consistent Match-Statistics: Standardizing Forensic Validation by Coupling Laboratory Specific Experimental Data with an In Silico DNA Pipeline

Dr. Catherine M. Grgicak, Dept of Anatomy & Neurobiology, Boston University School of Medicine

Samples containing low-copy or complex DNA mixtures are routinely encountered in operations. The signal acquired from these sample types are difficult to interpret as they do not always contain all of the genotypic information from each contributor, where the loss of genetic information is associated with sampling and detection effects. The present work focuses on developing a validation scheme to aid in mitigating the effects of the latter to produce high-fidelity electropherograms (EPGs) that can be effectively interpreted by all probabilistic genotyping systems.

To this end, we have devised a computational system, named RESOLVIt (Resolving Evidentiary Signal for Objective Laboratory Validation), that generates synthetic EPGs in a laboratory-specific manner. As an input to the system, a large number of single source profiles of known genotype are provided by the laboratory. From these data, the distribution of the peak heights at noise positions is modeled as a function of the starting template amount using a log-normal distribution. The electrophoresis sensitivity, which is used to generate the DNA height distribution, is also acquired from the single source experimental data procured from the laboratory. Other pertinent laboratory conditions, such as the number of PCR cycles, injection time, starting template mass, etc. are input parameters and are easily modified by the user.

Since RESOLVIt utilizes a simulation approach, which is based upon experimental data acquired from the laboratory, multifarious scenarios may be explored by each laboratory in a cost-effective manner. Metrics such as signal-to-noise resolution, false positive and false negative signal detection rates are used to select tenable laboratory conditions that result in high-fidelity signal in the single-copy regime. We demonstrate that the metrics acquired from simulation are consistent with experimental data obtained from two capillary electrophoresis platforms and various injection parameters. Once good resolution is obtained, analytical thresholds can be determined using detection error tradeoff analysis, if necessary.

Decreasing the limit of detection of the forensic process to one copy of DNA is a powerful mechanism by which to increase the information content on alleles from minor components of a mixture, which is particularly important for probabilistic system inference. By utilizing another fully continuous probabilistic system, CEESIt (Computational Evaluation of Evidentiary Signal), we demonstrate that if the forensic pipeline is engineered to produce high-fidelity EPG signal then the likelihood ratio (LR) of a true contributor increases and the probability that the LR of a randomly chosen person is greater than one decreases. CEESIt has been developed to not only compute the LR but also the probability that the LR is greater than one for millions of randomly chosen contributors, making it a powerful validation tool. This systematic, in-silico, laboratory-specific, computational-based approach to improve allele information content is, potentially, the first step towards standardization of the bio-analytical pipeline and DNA validation process across operational laboratories.

Date: Friday, October 13, 2017
Time: 11:00 a.m. to 12:00 p.m.
Location: AHC3– 205, MMC (Live)
Marine Sciences Building Room 105 (MSB-105) – BBC (via Polycom)