Research Commercialization

ValiDNA: Forensic DNA Software

Technology #2019-061

DNA collected from a crime scene or environment can be used to determine the likely sources that explain the data. DNA signal is complex as it contains representations of partial genomes from any number of unknown contributors, which is ultimately affected by the processing decisions made by the laboratory. As such, advanced computational systems that help laboratories choose optimal laboratory conditions and provide useful information on the number and type of contributors to the evidence are needed.

Summary

Rutgers researchers have developed a computational analysis invention and software (named ValiDNA) that solves these problems by generating synthetic DNA evidence for multifarious laboratory processes, estimating the number of contributors to a DNA sample and estimating the likelihoods of the DNA signal given a specific individual did or did not contribute to it. Features of ValiDNA include:

- Computation of the signal-to-noise resolution and allele/noise error detection rates for multiple laboratory scenarios through simulation such that laboratories may implement optimal laboratory and analysis conditions in a fast, cost-effective manner.
- Automatic filtering of artifacts such as pull-up, complex pull-up and minus A from electropherogram data.
- Determination of the probability of allele drop-in and drop-out.
- Determination of the likely number of contributors that comprise the DNA data,
- Determination of the likelihoods of the data given a specific individual did or did not contribute.

The ValiDNA software is comprised of the following five components:

- **ReSOLVIt**: Reports false positive and false negative detection rates for multifarious forensic DNA pipelines to allow forensic laboratories to optimize laboratory conditions by simulating forensically relevant samples at these conditions.
- **CleanIt**: Automated procedure to filter artifacts from forensically relevant DNA data, significantly decreasing pre-processing burdens on the analyst.
- **NOCIt**: Computes the posterior probability distribution on the number of contributors that make up the forensic DNA profile.
- **CEESIt**: Computes the likelihood ratio for the person-of-interest and millions of randomly generated persons.
- **CALLIt**: Parameterizes the models used in NOCIt and CEESIt

End to End Flow
All of the data are analyzed using a peak detection software of choice. Optimization of laboratory process is accomplished during Validation phase with ReSOLViIt. Once optimized laboratory conditions are determined, calibration data are garnered from single-source profiles of known genotype analyzed using the lowest possible signal threshold setting, and well-characterized artifacts, such as pull-up and minus A, are filtered with the CleanIt module and user-defined criteria. These calibration data are used to parameterize the models utilized by NOCIt. NOCIt determines the APP distribution on the NOC from data acquired from an unknown sample containing any number of contributors in any proportion. As with the calibration data, the STR data acquired from the environmental sample will undergo processing steps wherein peak detection and artifact filtering are completed. Unlike the calibration data, however, an analytical threshold, may be applied to the unknown, if desired. The data, containing information on the peak height (or signal intensity), allele call/size, are imported into NOCIt for evaluation. NOCIt outputs $P(N=n|E, H_d)$ for all $n$. Using the same models, CEESIt is then utilized to compute the full-likelihood ratio (LR) or LRs at different $n$ for the person-of-interest (POI). LR distributions for randomly generated genotypes, or other reporting statistics such as a p-value are also reported. For this pipeline, data from the calibration and unknown samples are expected to be acquired using the same DNA laboratory processing protocols (i.e., same STR assays, cycle numbers, and fragment analysis settings).

**Market Application:**

Aids forensic or ecological DNA experts or laboratories or any PCR-electrophoresis based laboratory interested in determining the number of contributors or likelihood ratio to a forensic or unknown sample

**Advantages:**

- Efficient, comprehensive and cost-effective for laboratory optimization and validation allowing a forensic laboratory to execute performance-based forensic validations by evaluating the impact on the end result.
- Automatic artifact filtering and model parameterization makes this method unique and effective
- Evaluation of the signal across all likely number of contributors without signal thresholds allows for full and complete evaluation of all the DNA data
Inventors

Catherine Grgicak

Catherine Grgicak, PhD is Henry Rutgers Chair and Associate Professor of Chemistry at Rutgers University – Camden specializing in forensic analysis and human identification. She earned her B.Sc. (Physical Sciences) and B.Ed. at the University of Windsor, her M.S.F.S. at the University of Alabama, Birmingham and her Ph.D. in Chemistry from the University of Ottawa in 2007.

Desmond Lun

Desmond Lun, PhD is a Professor in the Department of Computer Science at Rutgers University–Camden and the Department of Plant Biology at Rutgers University–New Brunswick. He received bachelor’s degrees in mathematics and computer engineering from the University of Melbourne, Australia in 2001, and S.M. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT) in 2002 and 2006, respectively.