Boston University School of Medicine
Biomedical Forensic Sciences

72 E Concord St. Rm R806B
Boston, Massachusetts 02118
www.bumc.bu.edu/gms/biomedforensic

Establishing Exclusion Criteria and the Significance of Inclusion for Complex Low-Template DNA Mixtures

FINAL SUMMARY OVERVIEW

June 30, 2018

Department of Justice, National Institute of Justice
Award Number: 2014-DN-BX-K026
(January 1, 2015 – Sept 30, 2018)

Principal Investigator(s):

Robin Cotton, PhD
Associate Professor
Program in Biomedical Forensic Sciences
Boston University School of Medicine
Boston, MA 02118
e-mail: rwcotton@bu.edu

Catherine Grgicak, PhD
Associate Professor
Department of Chemistry
Center for Computational and Integrative Biology
Rutgers University, Camden
Camden, NJ 08102
e-mail: c.grgicak@rutgers.edu

Contents

1. Purpose of the Project

The purpose of this work is to address the need to increase the knowledge and understanding associated with complex forensic DNA interpretation and to continue development of a novel approach to interpret low-template DNA samples containing many contributors. The specific aims were to: 1. Generate 2,400, well-defined, DNA autosomal-, mega-plex mixture profiles by varying the amount and ratio of cellular material in samples containing one- to five- contributors and using micromanipulation to sample the cells; 2. Refine statistical models that define stutter, allele dropout, baseline noise, allele peak height/area and confirm cross-kit and cross-platform compatibility for CEESIt – a continuous, freely available probabilistic system; 3. Develop a user-friendly probabilistic software system, CEESIt, that utilizes direct comparisons such that the match statistic (i.e., LRs) and operational conditions can be examined for relevance and optimization; 4. Update CEESIt to model degradation as well as differential degradation, and update the Graphical User Interface.

2. Project Design and Methods

2.1 Generate 2,400, well-defined, DNA mixture profiles with expanded STR amplification kits

We tested the viability of two MM-techniques to generate well-characterized mixtures: Laser Microdissection (LMD) and pico-pipetting. To test the viability of the LMD system, individual cells were dissected until 20 or 100 cells were added to a single well. Once dissected, DNA extraction using the LMD protocol provided with the Qiagen Investigator kit [1] ensued. In addition to LMD, we explored a pico-pipetting procedure, where the sample is neither dried, nor cut; rather, whole cells are pipetted into a small glass pipette tip ~75 µM in diameter by electro-osmotic flow. The cell is transferred directly to a well containing buffer and subjected to direct-PCR. Metrics of PCR recovery as per qPCR results acquired from the Quantifiler™ Trio assay, following the manufacturer's recommended protocol are reported.

We tested the extraction proficiency of four commercially available single-cell extraction procedures: *forensic*GEM® Extraction Kit (ForensicGem), DEPArray LysePrep Extraction Kit (Menarini Biosystems), DirectPCR Lysis solution (ThermoFisher) and PicoPure DNA Extraction Kit (ThermoFisher). Epithelial cells were suspended in TE Buffer and 150 µL of cell solution was aliquoted onto a glass slide for picopipetting. Each kit was tested using 102 single cell samples, consisting of 34 cells from three individuals. The total volume of the extraction mix used per sample for each kit was 5 µL. We leveraged our previous work and parameterized our *in silico* model, ReSOLVIt, with existing experimental data to quickly optimize amplification volumes and injection scenarios focusing on amplification volumes of 13 µL [2]. We acquired optimal signal detection rates when the DNA was amplified for 30 cycles using GlobalFiler® Amplification Kit (ThermoFisher) at half-reaction volumes (i.e., 13 µL), and fragment separation is accomplished on a 3500 Genetic Analyzer using 25 sec injections at 1.2 kV. These conditions were, therefore, used to generate the signal from the cells which were sampled using micromanipulation.

2.2 Refine the Models in CEESIt and Model degradation

From single-source samples, we characterize the signal amplitude and decay (A_c, B_c) for each dye color c, where A_c is the expected ***signal amplitude***, for color c, without degradation, and B_c is the ***decay factor***, which reflects the degradation of the sample for color c. We define the ***amplitude*** of the signal for a given locus, as the sum of all observed peaks at the locus and for a dye color that presents at least two loci of average size s_l, as an exponential regression curve of the form $x_l = A_c . e^{B_c . s}$ which has a unique solution (A_c, B_c).

The model includes four categories of peaks: 1) True Peaks, which are the result of the amplification of an allele that harbors a contributor; 2) Reverse Stutter; 3) Forward stutter, both caused by the strand slippage during replication; and 4) Noise Peaks which are all the peaks that

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

does not fit into any of the previous categories. For each peak category, we model its height using a Normal Gaussian density, and its frequency through a drop-out model. Drop-out of true peak appears, for example, when the DNA copies in the sample are too degraded to be amplified. We model the frequency of drop-out for true peaks and stutter peaks using an exponential decay.

Table 1: for each component, we indicate the probability distribution and its analytical form and the input (x) : $\mathbf{a_{ci}}.\mathbf{e^{b_{ci}s}}$ is the decayed amplitude of contributor *i*, at a color dye *c*, for an allele of size s; j_{N+1} and j_{N-1} refer to parent peak for reverse and forward stutter. Peaks models follow a normal density, and the frequencies of drop-out are modeled using an exponential decay. Noise drop-out parameter *a* is independent from the observed sample.

Model Component	PROBABILITY FUNCTION, PARAMETERS	MODEL INPUT
True Peaks	$\mathcal{N}\begin{pmatrix}\mu = u(x) = ax + b \\ \sigma = v(x) = cx + d\end{pmatrix}$ $\{a, b, c, d\}$	$x = \mathrm{a}_{ci}.e^{\mathrm{b}ci s}$
Reverse Stutter Peaks		$x = h_{N+1}$
Forward Stutter Peaks		$x = h_{N-1}$
Noise Peaks		$x = \mathrm{a}_{ci}.e^{\mathrm{b}ci s}$
Allele Drop-Out	$p(x) = ae^{-b.x}$ $\{a, b\}$	$x = \mathrm{a}_{ci}.e^{\mathrm{b}ci s}$
Reverse Stutter Drop-Out		$x = h_{N+1}$
Forward Stutter Drop-out		$x = h_{N-1}$
Noise Drop-Out	$p(x) = \boldsymbol{a}$	

CEESIt is built on a continuous mixture interpretation model that incorporates noise, stutter, stochastic PCR effects as well as random contributor levels as summarized in Table 1. Let E denote the evidence, i.e. the sample's electropherogram. Let H_1(g) denote the hypothesis that E arises from a contributor with genotype g in conjunction with a fixed number $k \geq 0$ of unknown contributors whose genotypes are selected randomly with given frequencies. Let H_2 denote the hypothesis that all $k+1$ contributors have genotypes selected at random with those given frequencies. Define the likelihood ratio for a specific genotype g to be

$$LR(g) = \frac{\Pr(E|H_1(g))}{\Pr(E|H_2)} \quad \text{(Equation 1)}$$

With p denoting a suspect's genotype and G denoting a randomly selected genotype, CEESIt computes approximations to $LR(p)$ and to the distribution of the random variable $LR(G)$. It

achieves this by computing an approximation of $\Pr(E|H_1(g))$, $\Pr(\widehat{E|H_1}(g))$, through Monte Carlo sampling. In addition, CEESIt generates a large number, *n*, of hypotheses $H_{2,1}$, $H_{2,2}$, …, $H_{2,n}$ where the genotypes of all *k+1* contributors are randomly chosen independently according to their frequencies in the population, excluding the genotype *p*. For each hypothesis $H_{2,i}$, it computes an approximation of $\Pr(E|H_{2,i})$, $\Pr(\widehat{E|H_{2,i}})$, through Monte Carlo Sampling. It then approximates $\Pr(E|H_2)$ as

$$\Pr(\widehat{E|H_2}) = \Pr(\widehat{E|H_1}(p))\Pr(G = p) + (1 - \Pr(G = p))\frac{1}{n}\sum_{i=1}^{n}\Pr(\widehat{E|H_{2,i}}). \quad \text{(Equation 2)}$$

CEESIt reports the approximate likelihood ratio for the suspect, $\text{LR(s)} = \Pr(\widehat{E|H_1}(p)) / \Pr(\widehat{E|H_2})$ and the Monte Carlo approximation to the distribution of $LR(G)$ obtained by $\Pr(\widehat{E|H_{2,i}}) / \Pr(\widehat{E|H_2})$ for $i = 1 \ldots n$.

2.3 Developing a User-Friendly GUI

Software validation was conducted in accordance with the General Principles of Software Validation, Version 2.0 by the Center for Devices and Radiological Health [3]. First, a risk assessment was performed to assign the criticality and complexity level of the software. We categorized CEESIt as critical and complex: critical because of its ability to substantially influence forensic DNA interpretation, statistical conclusions and the accuracy of the results; and complex because it contained many lines of code, complex algorithms and interconnected modules. Due to its classification CEESIt software validation testing included: 1) functional; 2) reliability; and 3) regression testing. Pre-determined acceptance criteria were recorded and are the expected responses from CEESIt based on the user requirements and test type. If the software output satisfied pre-determined requirements, the software functionality was categorized as "By Design." However, if CEESIt failed to meet acceptance criteria, remediation was required. Once the appropriate software modifications were made, a new distribution of CEESIt was released for

testing by the test team. Final release of the GUI (www.lftdi.com) was approved once all critical regression and functional tests were passed.

3. Data Analysis

3.1 *2,400, well-defined, DNA mixture profiles with expanded STR amplification kits*

No DNA signal was acquired from LMD cut individual cells. We, therefore, introduced and tested another method of micromanipulation; that of pico-pipetting. Figure 1 shows a representative electropherogram from a single cell extracted and amplified directly demonstrating we can achieve high-quality STR profiles with this technique.

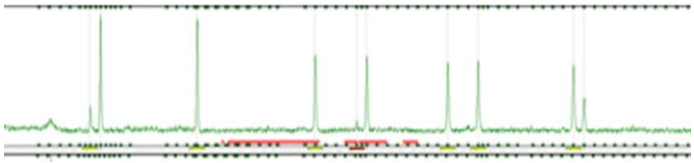

Figure 1. The green channel of an individual cell's DNA profile from picopetting coupled with a forenicGem lysis and GlobalFiler amplification.

Given the positive results of the pico-pipette technique, we explore DNA signal of four commercially available directPCR extraction chemistries. The height distribution of the heterozygous allelic peaks for two of the four extraction kits is shown in Figure 2. Across all kits, the peaks have heights are several hundred RFUs in magnitude, with a median height of 481, 408, 452 and 580 RFU for ForensicGem, LysePrep, DirectPCR and PicoPure, respectively. The heterozygous balance (Hb) and degradation effects between kits was also assessed[4, 5]. Degradation was evaluated by fitting an exponential. curve to the allele heights observed according to the

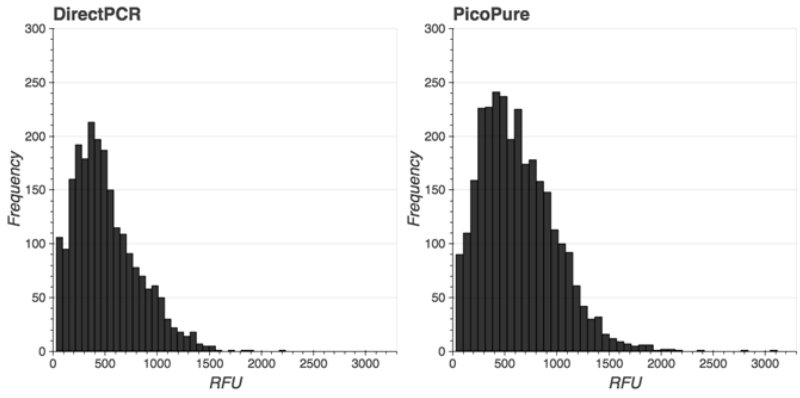

Figure 2. Peak height (RFU) distributions of STR peaks obtained for the four extraction kits.

model $y = ae^{bx}$. The 'b' parameter measures the level of degradation. No substantive differences in dropout rates, *Hb*, *b*, or peak heights between extraction kits were noted, suggesting the four kits contain viable chemistries to generate well-characterized DNA signal.

Using these data, we evaluate allelic dropout of samples from three people, Persons 01, 05 and 06, each of who have 34 heterozygous alleles. Thirty-four individual cells were analyzed per person for each of four extraction kits, giving a total of 4,624 heterozygous allelic positions per-person. Figure 3 plots a representative histogram of the number of alleles observed for Person 01's 136 single-cell profiles (blue histogram). Across the three persons, most of the profiles rendered were of 'good quality' where at least 75% of the heterozygous alleles were labeled, and the modes of the histograms are located at ca. 30-32 detected alleles. If allele dropout were independent, nearly all EPGs would result in partial profiles as the number of recovered alleles per profile would follow a Binomial distribution on 34 trials (red histogram). These data are inconsistent with the experimental results, demonstrating allele dropout rates are

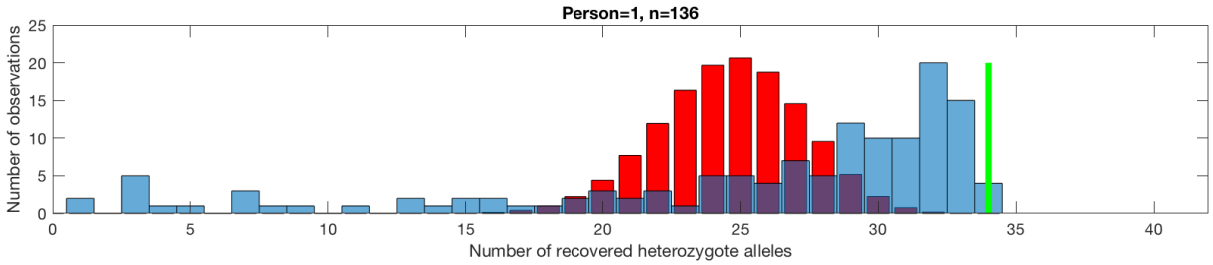

Figure 3. Histograms of the 'Number of recovered heterozygous alleles' from 136 individual cells from Persons 01. (■) Maximum number of recoverable alleles, 34, per EPG. (■) Histogram of number of alleles above an RFU of 30 per EPG fractionated by person tested. (■) Best-fit distribution of the number of recovered alleles if allele dropout was independent of the cell and locus.

not cell independent. Stutter Ratios (SRs) from our individually sampled cells were also assessed: At relatively large peak heights, i.e., >500, many stutter ratios were in excess of 15% - the expected SRs for high copy number samples[6]. For 2.15% of all measurements, however, the SR was greater than 1.

Table 2. Summary of the number of amplicons generated from pico-pipetting individual cells to generate well-characterized EPG signal from mixtures of 1- to 5- contributors.

No. of Contributors	No. of Genotype Combinations	Mixture Ratio	No of Amplifications
1	7	N/A	560
2	6	1:1 and 1:19	208
3	10	1:1:1 and 1:9:10 and 1:1:18	440
4	6	1:1:1 and 1:1:1:17 and 1:6:6:7	310
5	9	1:1:1:1:1 and 1:4:5:5:5 and 1:1:1:1:16	309
		Total:	1827

When cells, rather than DNA, are sampled (as is the case in directPCR), sampling and allele dropout effects cannot be assumed to be cell independent. Given these findings and our aim of constructing well-characterized signal, we generated 1827 electropherograms from individually sampled cells from cellular admixtures containing 1- to 5- contributors using pico-pipetting, PicoPure extraction and the GlobalFiler™ kit (Table 2).

We further augment the PROVEDIt database (www.lftdi.com) by amplifying 800 samples using traditional processes (i.e., aliquoting from a purified DNA extract).

Table 3. Mixed, degraded mixture sets proposed – PowerPlex Fusion 6C

No. of Contributors	Mixture Ratio	Masses (ng) of minor contributor	Level of degradation/ sloping	No. of Amplified Work-Products
1 (43 sets)	N/A	0.015, 0.031, 0.12, 0.5	0, 1, 2	516
2 (1 set)	1:1, 1:2, 1:4, 1:9	0.015, 0.031, 0.063, 0.125	0, 1, 2	32
3	1:1:1, 1:2:1, 1:4:1, 1:9:1, 1:2:2, 1:4:4, 1:9:9	0.015, 0.031, 0.063, 0.125	0, 1, 2	84
4	1:1:1:1, 1:1:2:1, 1:1:4:1, 1:1:9:1, 1:2:2:1, 1:4:4:1, 1:9:9:1	0.015, 0.031, 0.063, 0.125	0, 1, 2	84
5	1:1:1:1:1, 1:1:2:1:1, 1:1:4:1:1, 1:1:2:4:1, 1:1:2:9:1, 1:4:4:4:1, 1:9:9:9:1	0.015, 0.031, 0.063, 0.125	0, 1, 2	84
			Total:	800

3.2 Refine the Models in CEESIt and Model degradation

We tested CEESIt by exploring the effects of information content on the LR of a contributor and on the proportion of LRs greater than 1 for randomly sampled non-contributors. To evaluate the impact of signal content on these summary statistics we plot $\Pr(\widehat{LR(G)} > 1)$ when using an AT of 50 RFU and data generated from a 5s injection of 1-, 2- and 3-person samples on a 3130 Genetic Analyzer against $\Pr(\widehat{LR(G)} > 1)$ when the same samples were analyzed using an AT of 15 RFU (Blue and Green channels) or 20 RFU (Yellow and Red channels) and a 20 s injection (Figure 4a). When coupled with an AT of 50 RFU and the manufacturer's laboratory settings the detection error rates for noise and allele peaks were one due to high levels of allele drop-out when in the single-copy regime. In contrast, the 20 s injection time coupled with an AT of 15/20 RFU corresponded to detection error rates that did

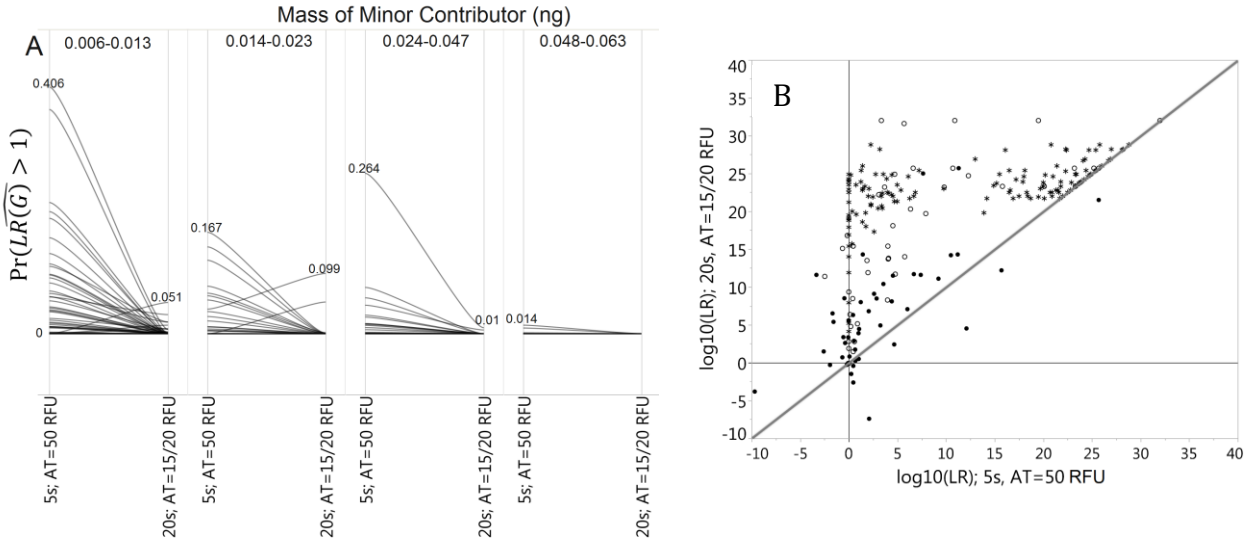

Figure 4. (A)Parallel plots of $\Pr(\widehat{LR(G)} > 1)$ for experimental 1-, 2- and 3-person samples injected for 5 s on the 3130 platform coupled with an AT of 50 RFU and injected for 20 s with an optimized AT of 15 and 20 for the blue/green and yellow/red channels, respectively. (B)Scatter plot of LRs for the true minor contributor of experimental (*) 1-, (○) 2- and (●) 3-person samples injected for 5 s with an AT of 50 RFU versus those injected for 20 s on the 3130 platform coupled with an AT of 15 and 20 for the blue/green and yellow/red channels, respectively. The x=y line is also shown. All samples were amplified with the Identifiler® Plus set of loci using 29 PCR cycles.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

not exceed 0.07. As a result, the probability of the LR > 1 for randomly generated genotypes decreases and the LRs for true contributors increases as additional allele information is imported into CEESIt (Figure 4), as expected.

3.3 Developing a User-friendly GUI

Most of the software tests types used to test the GUI were positive functional tests. Specifically, 68%, 1%, 13%, 2% and 16% were positive, negative, fuzz, boundary and regression tests, respectively. Over the course of 3 months of GUI testing and development the software acceptance rate increase from ca. 40% to 100%, demonstrating the value of implementing structured software testing procedures during development. In addition, batch-processing functionality was introduced and was confirmed to work with at least 20 samples in a batch. The upper bound on the number of contributors' assumption was 5-persons. The tests included data generated by two capillary electrophoresis platforms; data from three STR multiplex kits; multiple population groups; multiple processors/computers and operating systems; as well as different users.

4. List of Scholarly Products

1 NIJ Symposium at the American Academy of Forensic Sciences Conference – Catherine M. Grgicak. *"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"*, (Feb 2018, Seattle, WA)

2 43rd Northeastern Association of Forensic Scientists – Catherine M. Grgicak. *Production of High-Fidelity Electropherograms Results in Improved and Consistent Match-Statistics: Standardizing Forensic Validation.* (Oct 2017, Pocono Manor, PA)

3 43rd Northeastern Association of Forensic Scientists – Lauren Alfonse and Catherine M. Grgicak. *CleanIt: An Automated Procedure for Filtering Electropherogram Artifacts.* (Oct 2017, Pocono Manor, PA)

4 43rd Northeastern Association of Forensic Scientists – Jennifer Sheehan and Catherine M. Grgicak. *Characterizing Double-Back Stutter in Low- to Multi-Copy Number Regimes in Forensically Relevant STR Loci.* (October 2017, Pocono Manor, PA)

5 International Symposium on Forensic Science Error Management – Harish Swaminathan and Catherine M. Grgicak. *Parameterization of an in silico DNA Pipeline with Laboratory-Specific Experimental Data Allows for Efficient Validation of the DNA Analysis Process* (July 2017, Gaithersburg, MD)

6 L. E. Alfonse, A. D. Garrett, D. Lun, K. R. Duffy & C. Grgicak. *A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt.* Forensic Science International: Genetics, 32, 62-70, (2018)

7 K. Peters, H. Swaminathan, K. R. Duffy, D. Lun, J. Sheehan, & C. Grgicak. *Production of high-fidelity electropherograms results in improved and consistent DNA interpretation: standardizing the forensic validation process*. Forensic Science International: Genetics, 31, 160-170, (2017).

8 K.R. Duffy, Neil Gurram, K.C. Peters, G. Wellner, C.M. Grgicak. *Exploring Forensic STR Signal in the Single- and Multi-Copy Number Regimes: Deductions from an In Silico Model of the Entire DNA Laboratory Process.* Electrophoresis**,** 38(6), 855–868, (2017)

9 27th International Symposium on Human Identification – Lauren E. Alfonse, Amanda D. Garrett, Harish Swaminathan, Kelsey C. Peters, Genevieve Wellner, Xia Yearwood-Garcia, Lauren M. Taranow, Jennifer Sheehan, Sarah E. Norsworthy, Ullrich Mönich, Desmond S. Lun, Ken R. Duffy, Muriel Médard, Robin W. Cotton, Catherine M. Grgicak. *The Development and Release of a Collection of Computation Tools and a Large-Scale Empirical Data Set for Validation: The PROVEDIt Initiative.* (Sept 2016)

10 1st Gordon Research Conference: Forensic Analysis of Human DNA. *Forensic DNA Research, Validation and Pedagogy Using Empirical and Simulated Data: Understanding the Behavior of Mixtures* Waterville Valley, NH (June 2016)

5. Implications to Criminal Justice Policy and Practice

We have upgraded a continuous method to compute the LR and LR distribution based on modeling of the peak heights by simulation of genotypes based on allele frequencies, while modifying a previously developed full mechanistic model of the forensic laboratory process for purposes of forensic validation. CEESIt behaves as expected in that we show that if the LR for a true contributor is, in general, larger and the probability that the LR determined for a non-contributor is greater than one is reduced when more information is imported into CEESIt. In addition, we implemented a structured software testing process and results suggest that CEESIt v 3.1 is fit for its intended and is robust.

References

[1] Qiagen Inc., QIAamp(R) DNA Investigator Handbook, Venlo, 2010.
[2] K.C. Peters, H. Swaminathan, J. Sheehan, K.R. Duffy, D.S. Lun, C.M. Grgicak, Production of high-fidelity electropherograms results in improved and consistent DNA interpretation: Standardizing the forensic validation process, Forensic Sci. Int. Genetics 31 (2017) 160-170.
[3] C.f.D.a.R. Health, General Principles of Software Validation; Final Guidance for Industry and FDA Staff, 2002.
[4] L.E. Alfonse, A.D. Garrett, D.S. Lun, K.R. Duffy, C.M. Grgicak, A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt, Forensic Sci. Int. Genetics 32 (2018) 62-70.
[5] H. Kelly, J.-A. Bright, J.M. Curran, J. Buckleton, Modelling heterozygote balance in forensic DNA profiles, Forensic Science International: Genetics 6 (2012) 729-734.
[6] Life Technologies Corporation, GlobalFiler PCR Amplification Kit User Guide, Carlsbad, 2016.