NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Interpretation of Complex DNA Mixtures

Michael D Coble, Ph.D.

November 9, 2016

FORENSICS @ NIST

#NISTForensics

Official Disclaimer

The opinions and assertions contained herein are solely those of the author and are not to be construed as official or as views of the U.S. Department of Commerce.

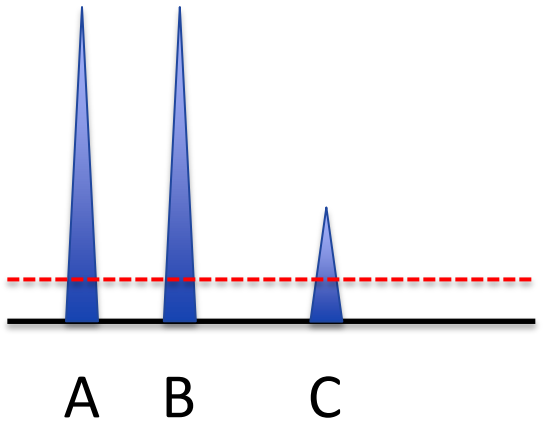
Commercial equipment, instruments, software, or materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the U.S. Department of Commerce, nor does it imply that any of the materials, instruments, software or equipment identified are necessarily the best available for the purpose.

Current DNA Interpretation Methods

Threshold-based Interpretation

1. Random Man Not Excluded (CPI or CPE)
   - Considers **all** possible genotype combinations

$$(f_a + f_b + f_c)^2$$



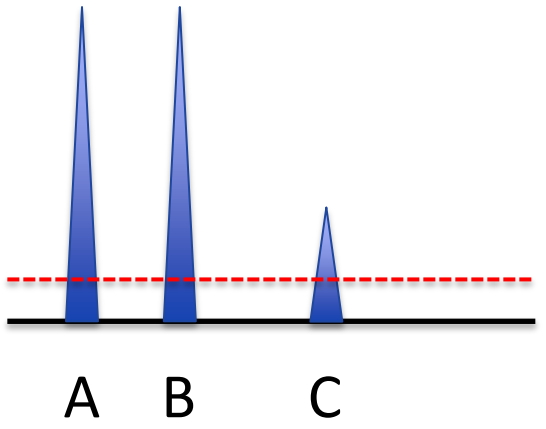
AA and BC - possible

Current DNA Interpretation Methods

Threshold-based Interpretation

1. Random Man Not Excluded (CPI or CPE)
   - Considers **all** possible genotype combinations

2. Modified Random Match Probability (mRMP)
3. Likelihood Ratio (LR)
   - assumptions to the number of contributors and other parameters (PHR, Mixture Ratios) can be used to restrict combinations



$$2f_a f_c + 2f_b f_c + f_c^2$$

If AB is the complainant, then POI = AC or BC or CC

Why Change?!?

(1) Drop-out – when alleles may be missing from the profile ('dropped' below the AT).

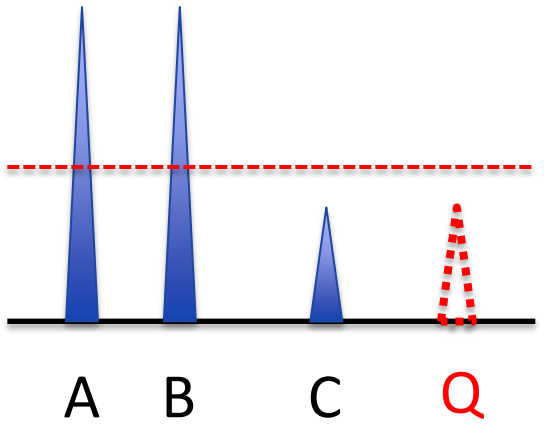
(2) Alleles that are between the Analytical Threshold and the Stochastic Threshold.

Current DNA Interpretation Methods

Threshold-based Interpretation

1. Random Man Not Excluded (CPI or CPE)
   - Considers **all** possible genotype combinations

$$(f_a + f_b + f_c + f_q)^2$$



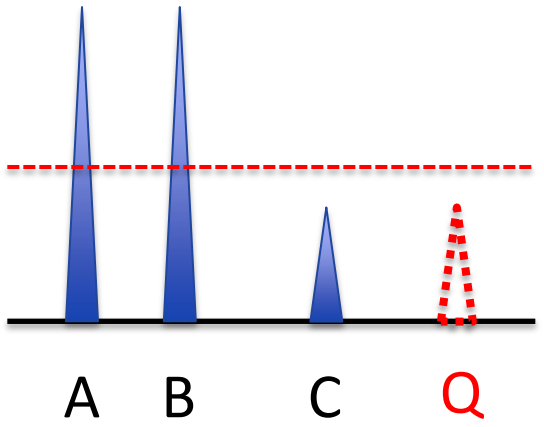
Any genotype is possible

Current DNA Interpretation Methods

Threshold-based Interpretation

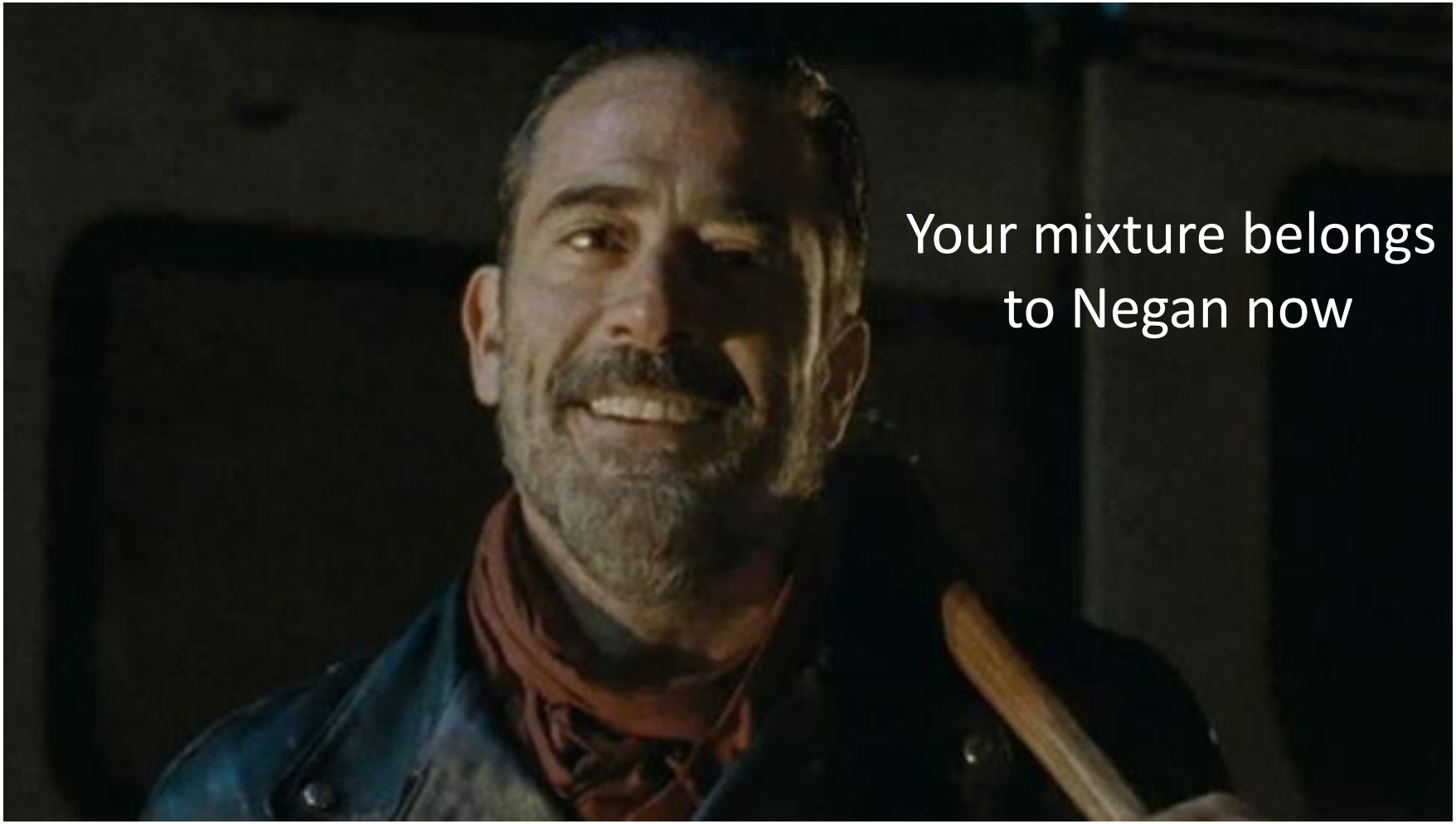
1. Random Man Not Excluded (CPI or CPE)
   - Considers **all** possible genotype combinations

$$(f_a + f_b + f_c + f_q)^2$$



Any genotype is possible

Current DNA Interpretation Methods

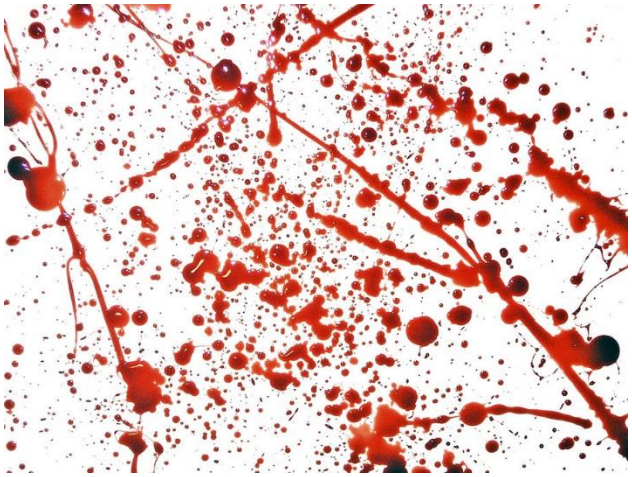
1. 

Current DNA Interpretation Methods

Threshold-based Interpretation

1. Random Man Not Excluded (CPI or CPE)
   - Considers **all** possible genotype combinations

$$(f_a + f_b + f_c + f_q)^2$$



Any genotype is possible

Current DNA Interpretation Methods

Threshold-based Interpretation

1. Random Man Not Excluded (CPI or CPE)
   - Considers **all** possible genotype combinations

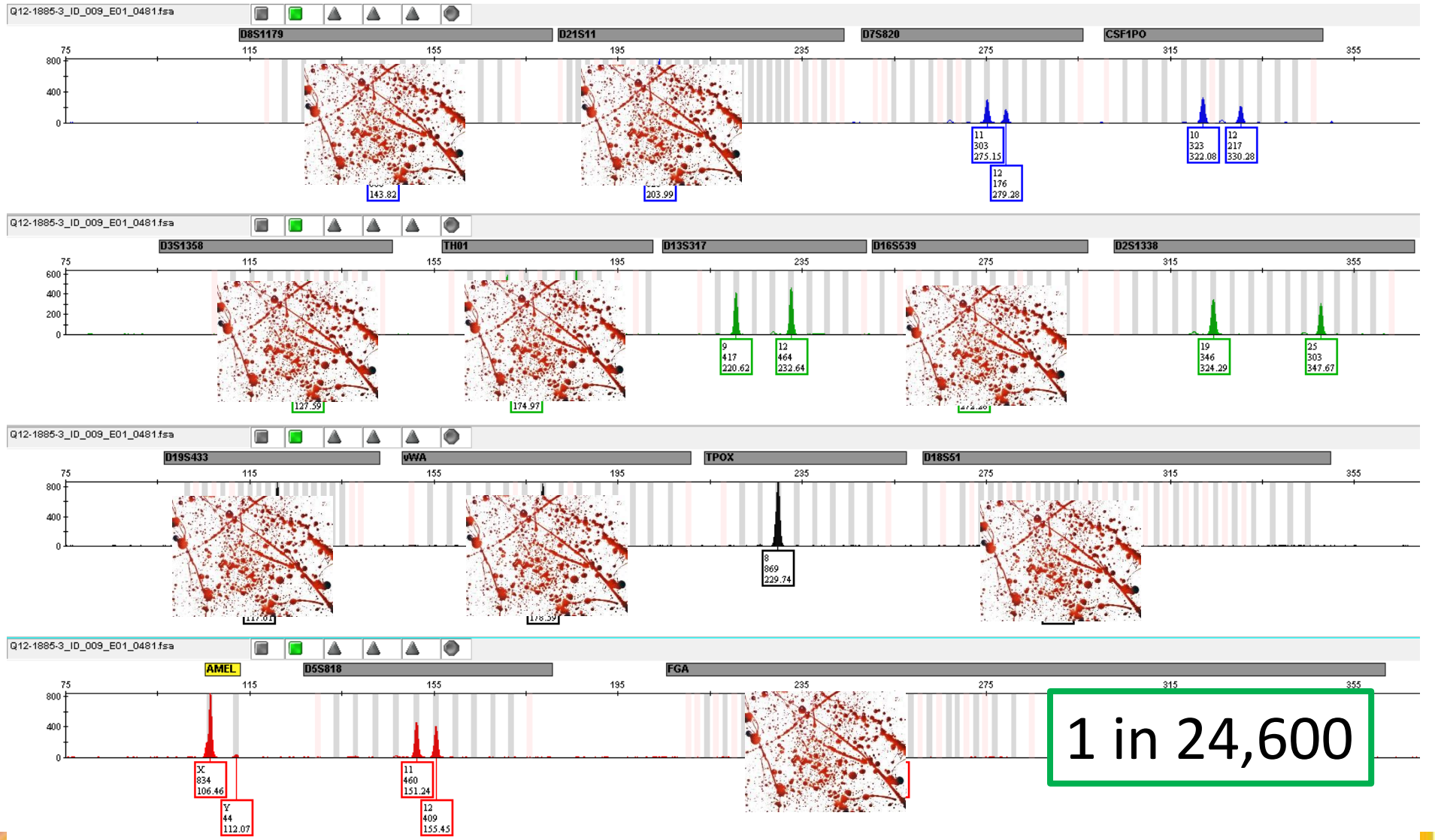
$$(f_a + f_b + f_c + f_q)^2$$

This locus is no longer available for statistics

Any genotype is possible

CPI

NIST
National Institute of Standards and Technology
U.S. Department of Commerce



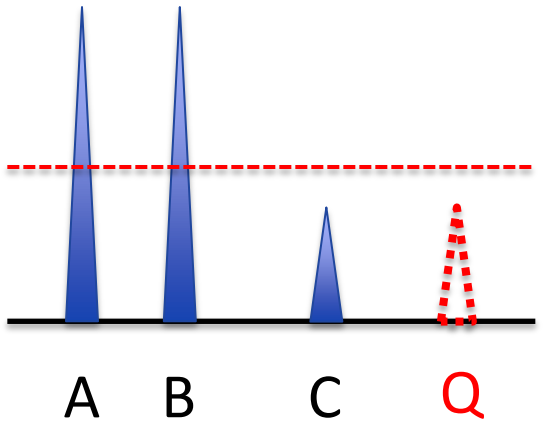
1 in 24,600

Current DNA Interpretation Methods

Threshold-based Interpretation

1. Random Man Not Excluded (CPI or CPE)
   - Considers **all** possible genotype combinations

2. Modified Random Match Probability (mRMP)
3. Likelihood Ratio (LR)
   - assumptions to the number of contributors and other parameters (PHR, Mixture Ratios) can be used to restrict combinations

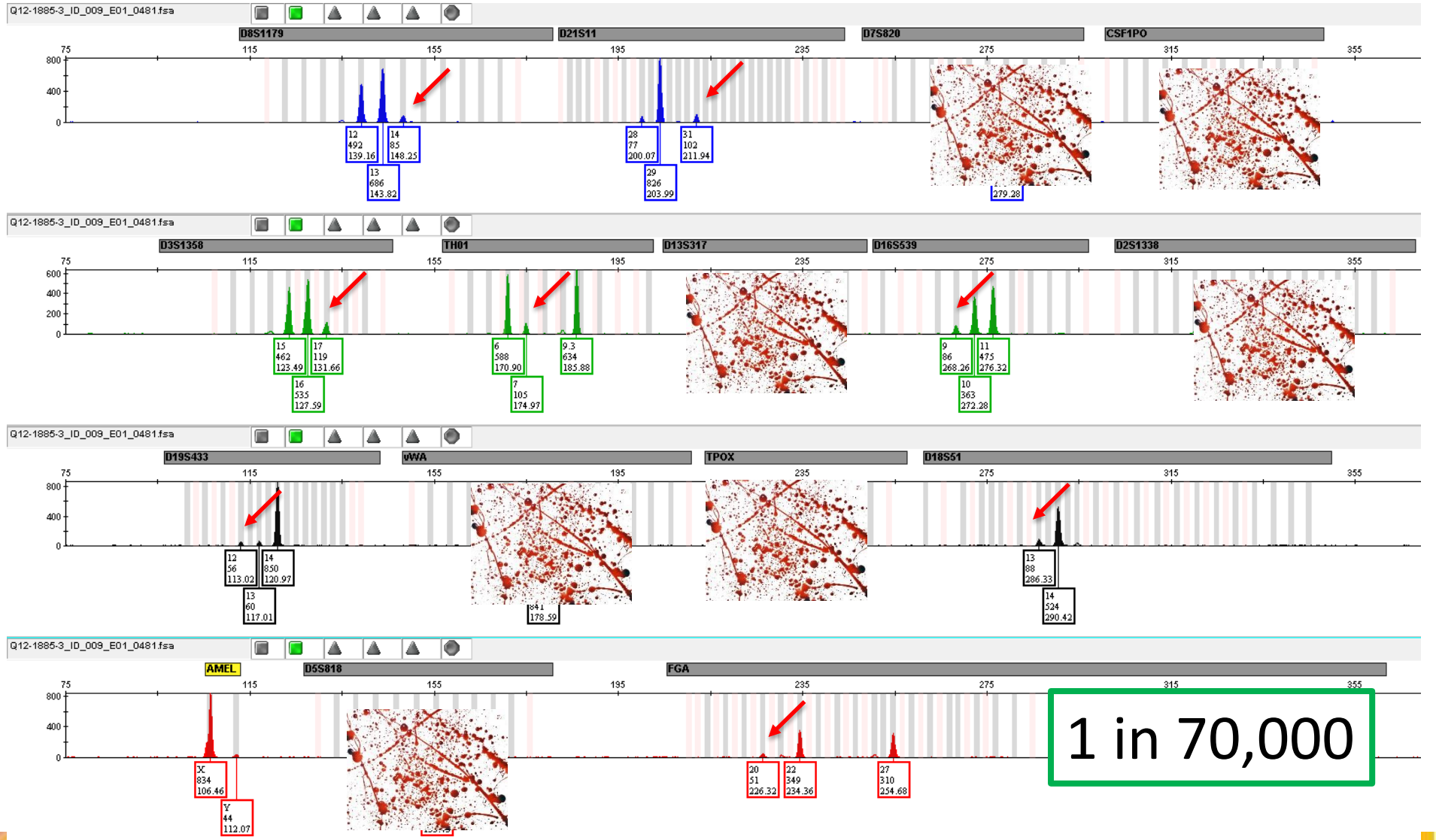


If AB is the complainant, then POI = CQ

$$2p = 2f_c$$

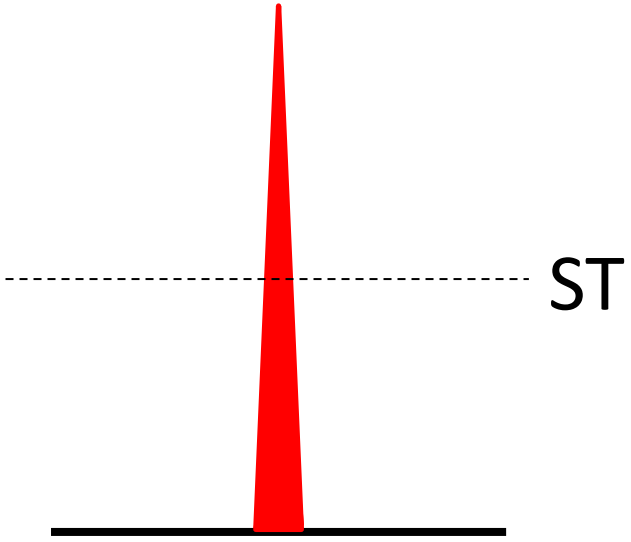
mRMP/LR

NIST
National Institute of Standards and Technology
U.S. Department of Commerce



1 in 70,000

The "2p" Rule



LR = 100

Stain = CC

POI = CC

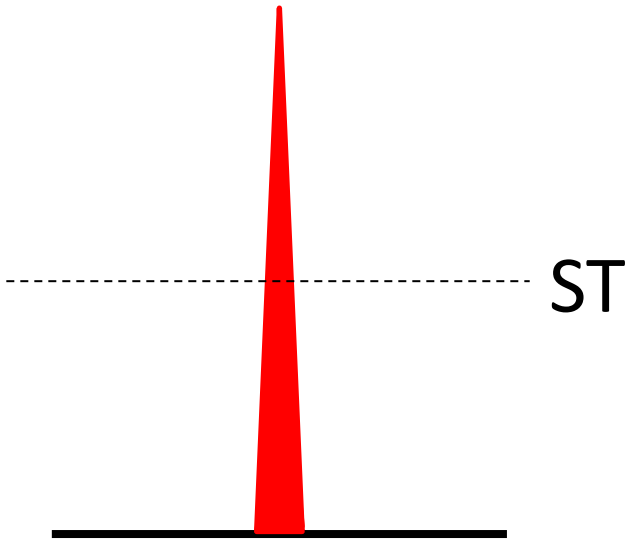
Let

$f_c = 0.10$

$1/p^2 = 100$

$1/2p = 5$

The “2p” Rule



Exclusion

Stain = CC

POI = AC

Let

$f_c = 0.10$

$1/p^2 = 100$

$1/2p = 5$

The Motivation for Change

- STR kits and CE instruments have become more sensitive than in the past.
- Evidence submitted to the lab has moved from predominately high quality/quantity sources to more trace profiles.

- More mixtures and increased stochastic effects

Probabilistic Genotyping

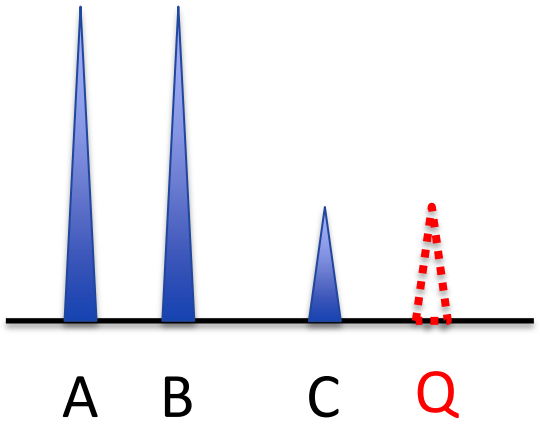
- "The use of biological modeling, statistical theory, computer algorithms, and probability distributions to calculate likelihood ratios (LRs) and/or infer genotypes for the DNA typing results of forensic samples."

SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems

Current DNA Interpretation Methods

Non *Threshold-based Interpretation*

1. Discrete Models of Interpretation
   - Considers **only** the alleles present
   - Peaks heights are ignored
   - Uses a Probability of dropout $Pr(D_{out})$ to account for missing alleles and a Pr (D_{in}) for any spurious alleles
   - All possible genotypes are considered



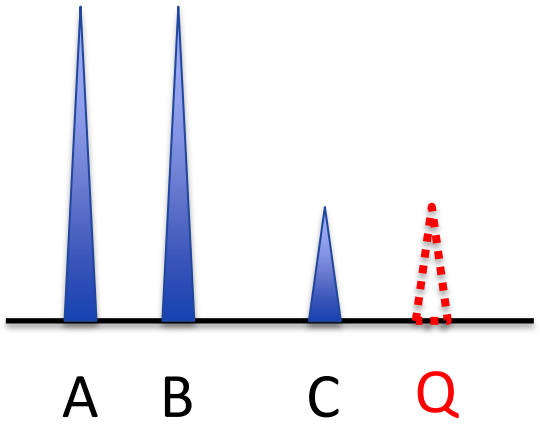
A = Yes
B = Yes
C = Yes

$Pr(D_{out}) = 0.15$
$Pr\ (D_{in}) = 0.01$

Current DNA Interpretation Methods

Non *Threshold-based Interpretation*

2. Continuous Models of Interpretation
   - Mathematical modeling of the profile to determine optimal genotypes
   - Peaks heights, mixture ratios, stutter, etc… are considered.
   - Drop out is modeled (not 'physically entered')
   - May use simulations via MCMC (not all programs!)



AC BC
CC CQ

Benefits of Probabilistic Genotyping

Electrophoresis 2014, *35*, 3125–3133

Todd W. Bille[1]
Steven M. Weitz[1]
Michael D. Coble[2]
John Buckleton[3]
Jo-Anne Bright[3]

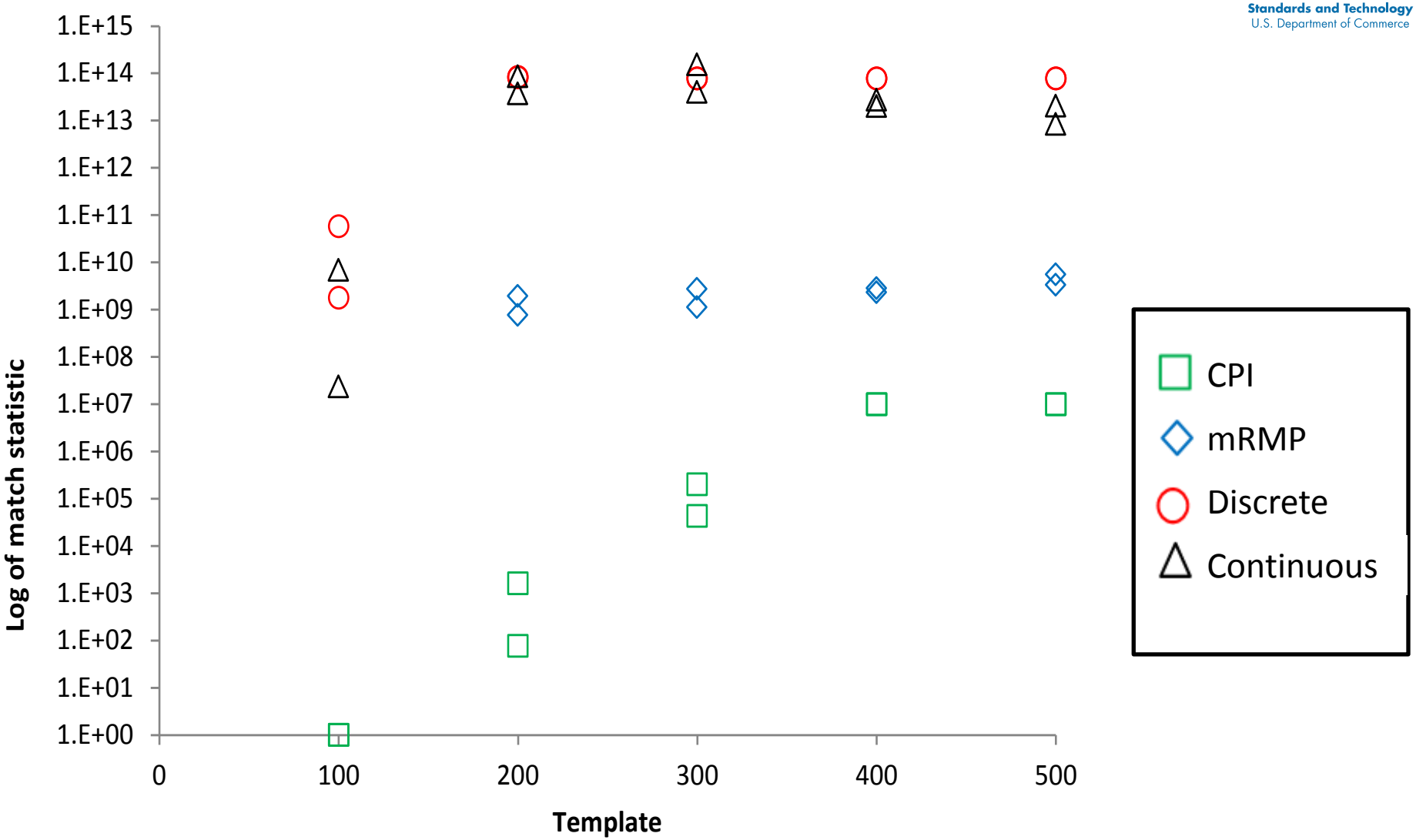
[1]Bureau of Alcohol, Tobacco, Firearms and Explosives,

Research Article

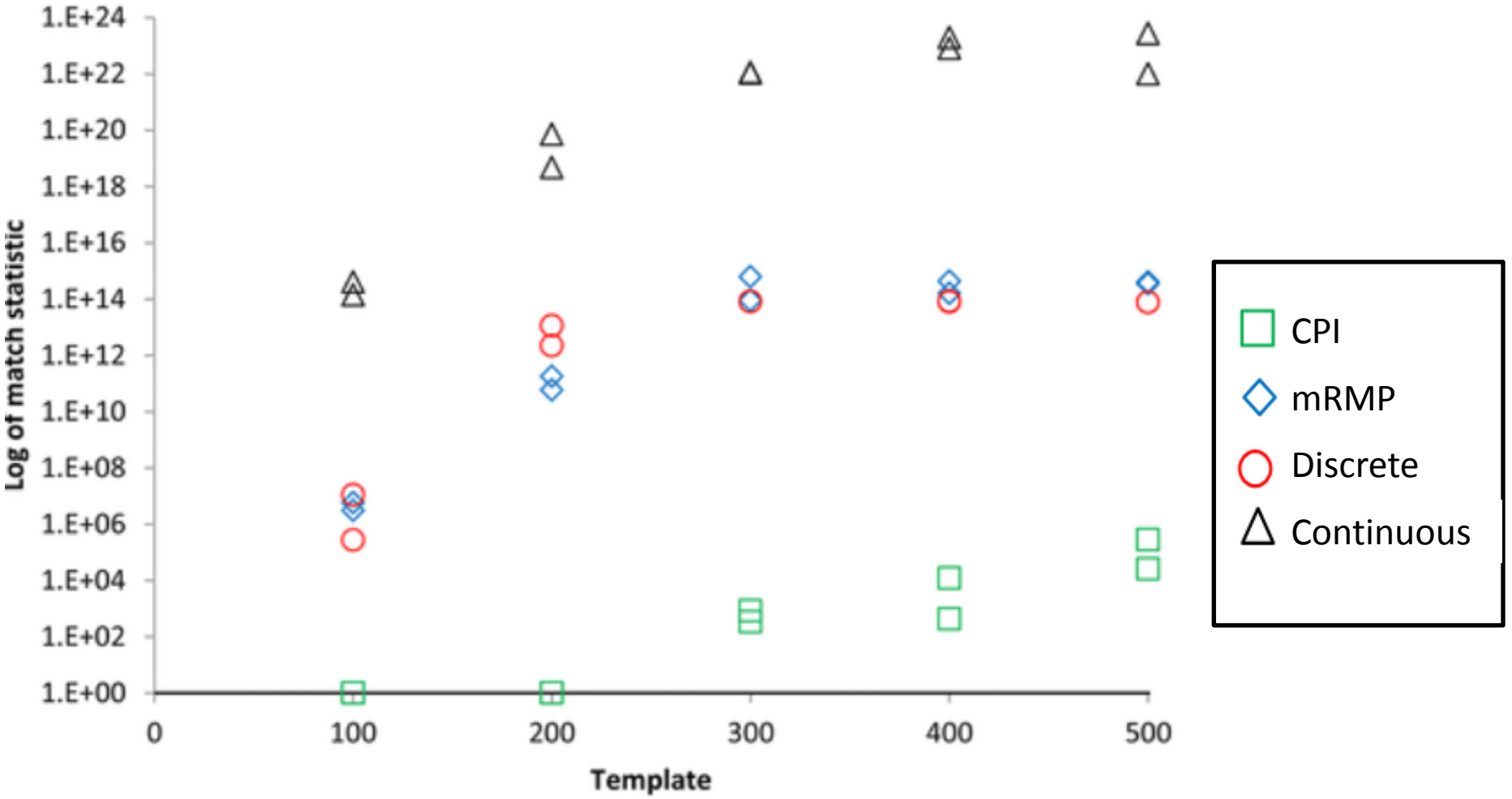
Comparison of the performance of different models for the interpretation of low level mixed DNA profiles

2 person mixtures at 5 different quantities and 5 different mixture ratios

CPI RMP (2P) Discrete Continuous



1:1 Mixture Ratio



3:1 Mixture Ratio

Landscape Study of DNA Mixture Interpretation Software

July 2015

11 PG Software Profiled

Principle Investigator:
Jeri Ropero-Miller
FTCoE Director
jerimiller@rti.org

Technical Contacts:
Patricia Melton
pmelton@rti.org

Lyndsie Ferrara
schantzl@duq.edu

Jonas Hall
jonashall@rti.org

https://www.forensiccoe.org/Our-Impact/Advancing-Technology/Reports/Demystifying-MIST-Landscape-Report-for-DNA-Mixture-Interpretation-Software-Tools

And…

EuroForMix

An open-source software for statistical DNA interpretation

likeLTD v6.1: an illustrative analysis, explanation of the model, results of validation tests and version history

David J. Balding, Christopher D. Steele
UCL Genetics Institute
Darwin Building, Gower Street
London WC1E 6BT
d.balding@ucl.ac.uk

June 5, 2016

Validation and Guidance

Scientific Working Group on
DNA Analysis Methods

Guidelines for the Validation of
Probabilistic Genotyping
Systems

Guidance for Developmental and Internal Validation
29 numbered recommendations

Validation and Guidance

Forensic Science International: Genetics 25 (2016) 191–197

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Research paper

DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications

M.D. Coble[a,*], J. Buckleton[b,c], J.M. Butler[d], T. Egeland[e], R. Fimmers[f], P. Gill[g,h], L. Gusmão[i,j,k], B. Guttman[l], M. Krawczak[m], N. Morling[n], W. Parson[o,p], N. Pinto[j,k,q,r], P.M. Schneider[s], S.T. Sherry[t], S. Willuweit[u], M. Prinz[v]

9 recommendations for software developers

7 recommendations for end-users

Validation and Guidance

- Other resources – references in the Mixture Software Landscape study
- Recent publications
- NIST validation page on STRbase?

Validation Resources (NIST)

Validation Information to Aid Forensic DNA Laboratories

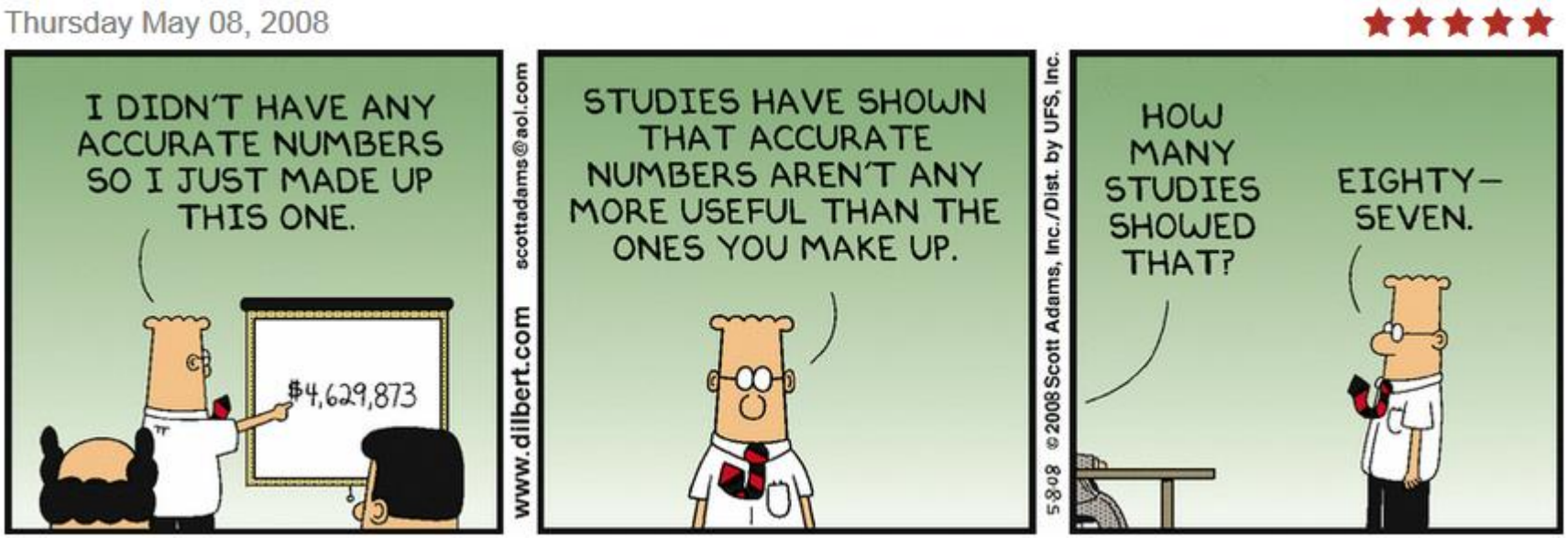
Validation Summary Sheets

We are initiating an effort to catalog and summarize validation studies that have been published order to aid current and future validation efforts by forensic DNA laboratories. These validation s summarized."

Below is listed a compilation of references to various validation studies conducted using comme **on the hyperlink to access a specific Validation Summary Sheet** *(note that not all validation summ*

Kit, Assay, or Instrument	Reference
PowerPlex Y	Krenke *et al.* (2005)
Profiler Plus	Frank *et al.* (2001), LaFountain *et al.* (2001), Tomsey *et al.* (2001), Holt *et al.* (2002), Fregeau *et al.* (2003), Buse *et al.* (2003), Wallin *et al.* (2002), Pawlowski *et al.* (2000), Moretti *et al.* (2001)
mtDNA minisequencing	Morley *et al.* (1999)
TrueAllele software	Kadash *et al.* (2004)

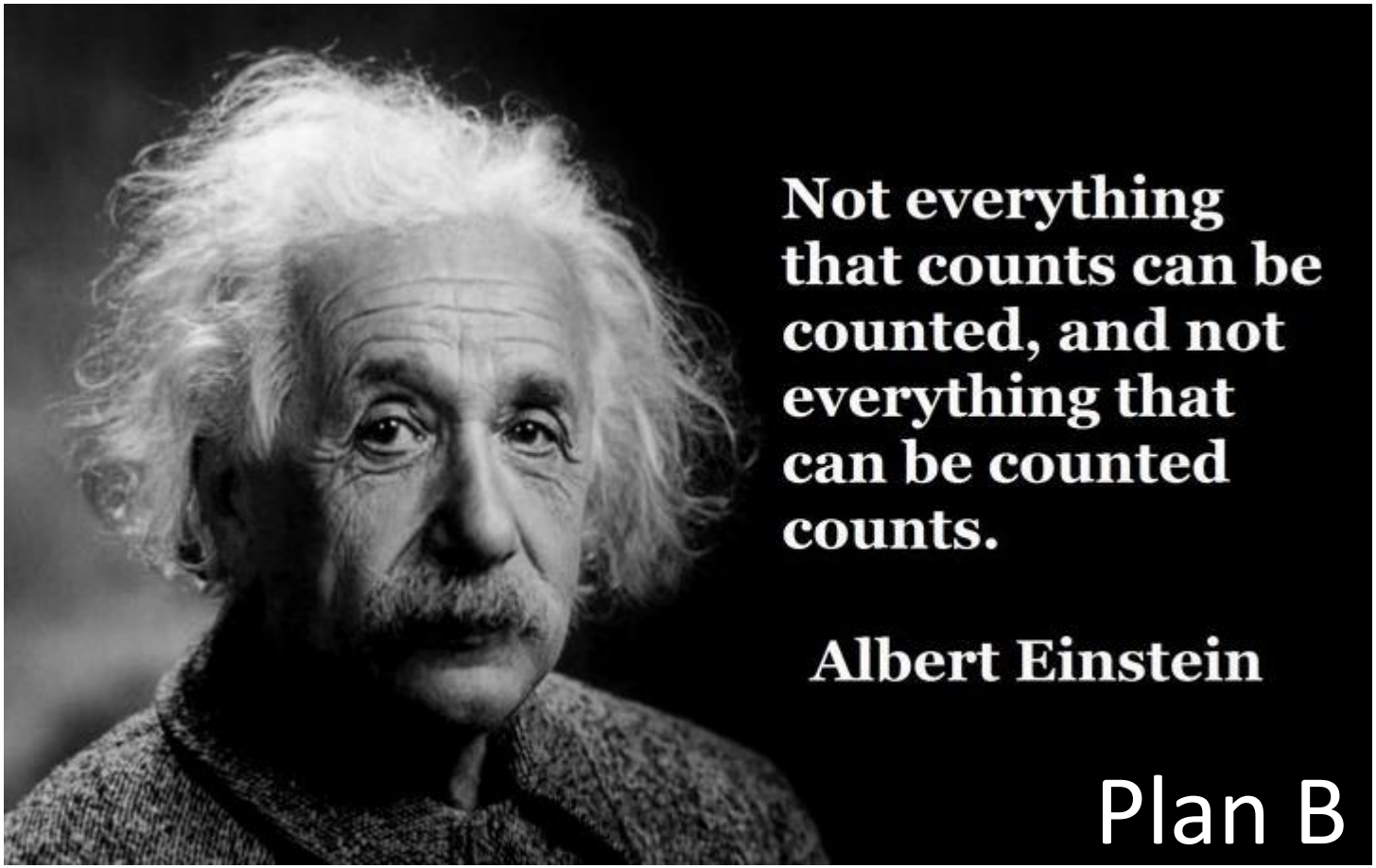
Validation Consensus Plan Ahead!



Plan A

Slide courtesy of Robin Cotton

Validation Consensus
Plan Ahead!



Slide courtesy of Robin Cotton

Summary

- Probabilistic Methods make better use of the data than RMNE or the binary LR with 2p.
- The goal of the software programs should not be to simply "get bigger numbers" but to understand the details of these approaches and not treat the software as a "black box."
- Know your models!!
- Understanding and properly using these software programs will become evident from a well planned and executed validation study.

Thank you!

Collaborators

Applied Genetics Group

Todd Bille & Steven Weitz (ATF)
Jo Bright (ESR)
John Buckleton (ESR and NIST)
Robin Cotton (BU)
Charlotte Word (Consultant)
John Butler (NIST SPO)

Funding

– NIST Special Programs Office (Forensic DNA)

michael.coble@nist.gov