BODE Conference
13 February 2023
Phoenix, AZ

LFTDI

LABORATORY FOR FORENSIC TECHNOLOGY DEVELOPMENT & INTEGRATION

The probability of the number of contributors given forensic DNA data

Catherine M. Grgicak

RUTGERS UNIVERSITY | CAMDEN

BOSTON UNIVERSITY

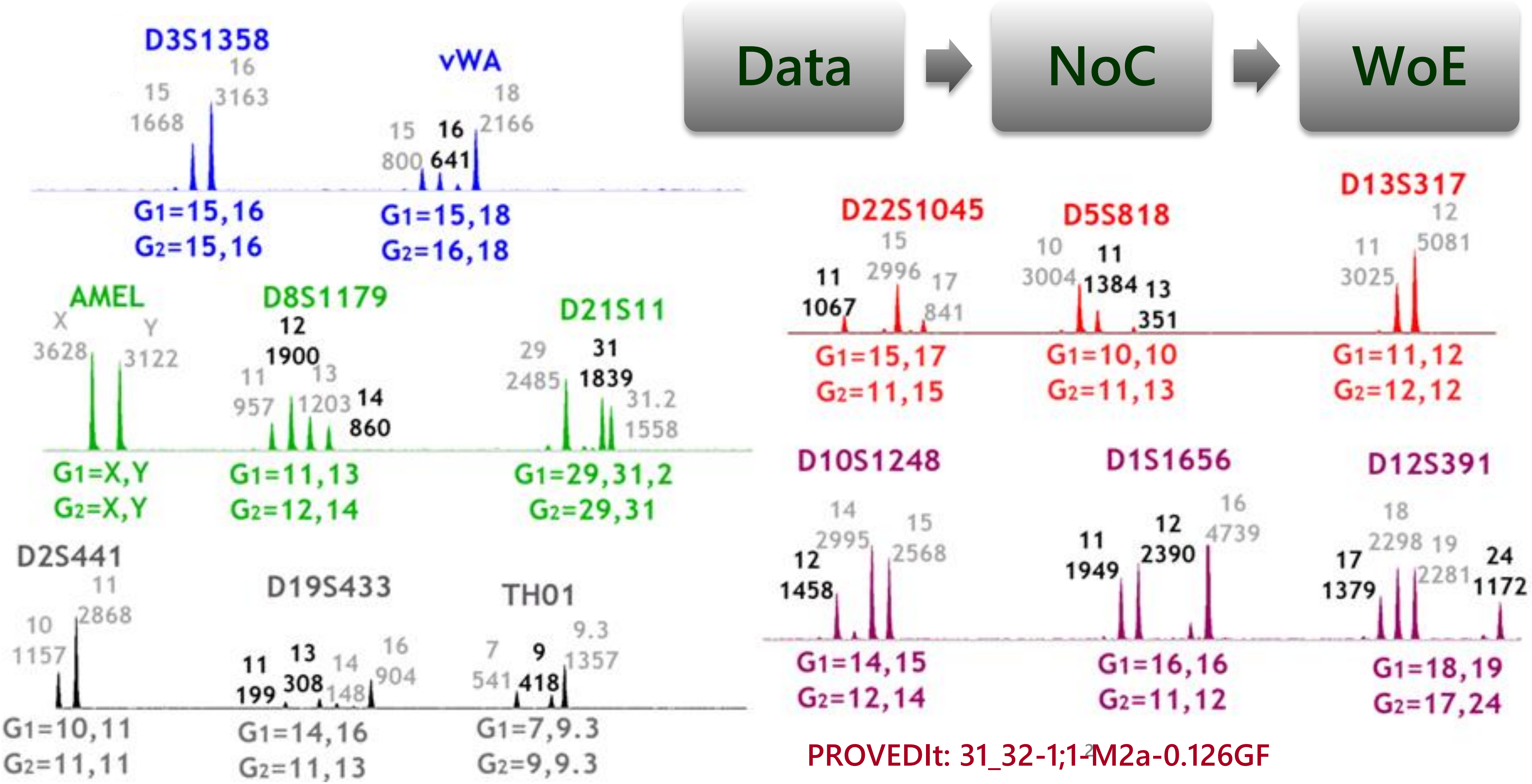
R CAMDEN

Maynooth University National University of Ireland Maynooth

Massachusetts Institute of Technology

SOFTGENETICS®

DNA Mixtures and Number of Contributors (NoC)



PROVEDIt: 31_32-1;1-M2a-0.126GF

NOCIt Gives $P(N=n|H_D,I,E)$ for all N

Why $P(N=n|H_d,I,E)$?

The LR is the weighted average of the n-specific LRs, for all n

Contents lists available at ScienceDirect
Forensic Science International: Genetics
journal homepage: www.elsevier.com/locate/fsigen

Research paper

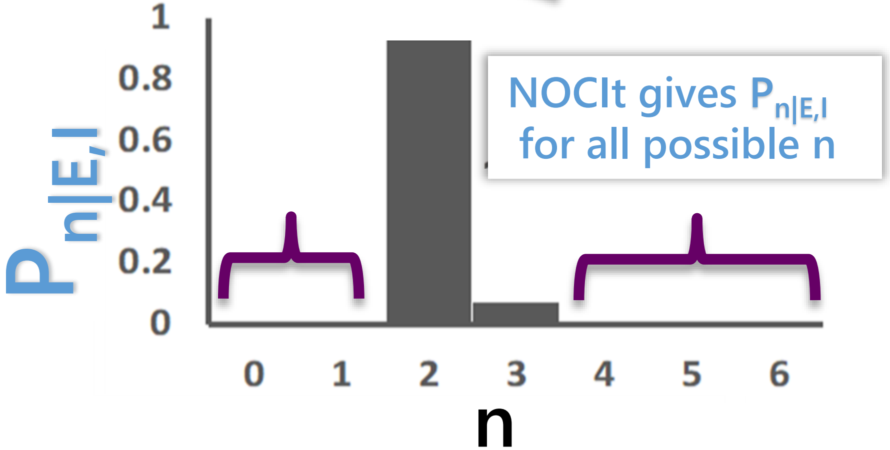
Contributors are a nuisance (parameter) for DNA mixture evidence evaluation

K. Slooten[a,*], A. Caliebe[b]

$$LR = \sum_n \left(\frac{P(E|H_p, N=n, I)}{P(E|H_d, N=n, I)} P(N=n|H_d, I, E) \right)$$

$LR^{E|n,I}$ $P_{n|E,I}$ NOCIt

$$LR = LR^{E|n=1,I}P_{n=1|E,I} + LR^{E|n=2,I}P_{n=2|E,I} + LR^{E|n=3,I}P_{n=3|E,I} + LR^{E|n=4,I}P_{n=4|E,I} + LR^{E|n=5,I}P_{n=5|E,I} + \ldots$$



NOCIt, therefore, meets 2 aims:

- Narrows *n* ranges by informing what n are associated with negligible $P_{n|E,I}$
- Supports a process that does not apply default *n* or automatic $P_{n|E,I}$

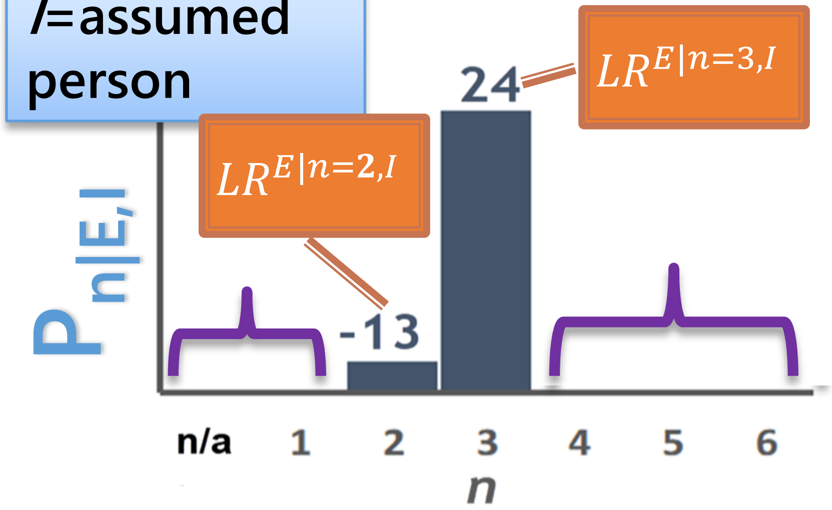
NOCIt Determines $P(N=n|H_D,I,E)$

Why $P(N=n|H_d,I,E)$ aka $P_{n|E,I}$?

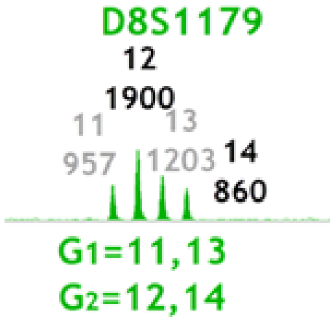
Contexts between $LR^{E|n,I}$ and $P_{n|E,I}$ are consistent

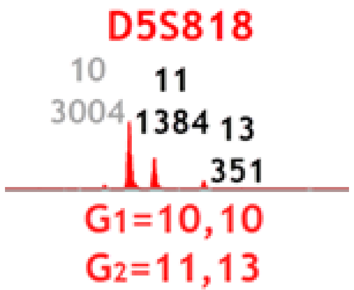
$$\mathbf{LR} = LR^{E|n=1,I}P_{n=1|E,I} + LR^{E|n=2,I}P_{n=2|E,I} + LR^{E|n=3,I}P_{n=3|E,I} + LR^{E|n=4,I}P_{n=4|E,I} + LR^{E|n=5,I}P_{n=5|E,I} + \ldots$$

I=assumed person



$$\mathbf{LR} = LR^{E|n=2,I}P_{n=2|E,I} + LR^{E|n=3,I}P_{n=3|E,I}$$
$$= 10^{-13}0.1 + 10^{24}0.9$$
$$= 10^{24}$$





NOCIt, therefore, meets a 3rd aim:

- Determines $P_{n|E,I}$ using the *same context* as assigned to $LR^{E|n,I}$

NOCIT DETERMINES $P(N=n|H_D,I,E)$

What is $P(N=n|H_d,I,E)$?

It is the posterior probability of n contributors given the data and a context

Probability of the data given n randomers

Probability that n people contributed (before seeing data – PRIOR)

$$P(N=n|H_d,I,E) = \frac{P(E|H_d,I,N=n)\cdot P(N=n|H_d,I)}{\sum_n P(E|H_d,I,N=n)\cdot P(N=n|H_d,I)}$$

Probability of n contributors after considering the data - POSTERIOR

For ease, drop I and H_d going forward

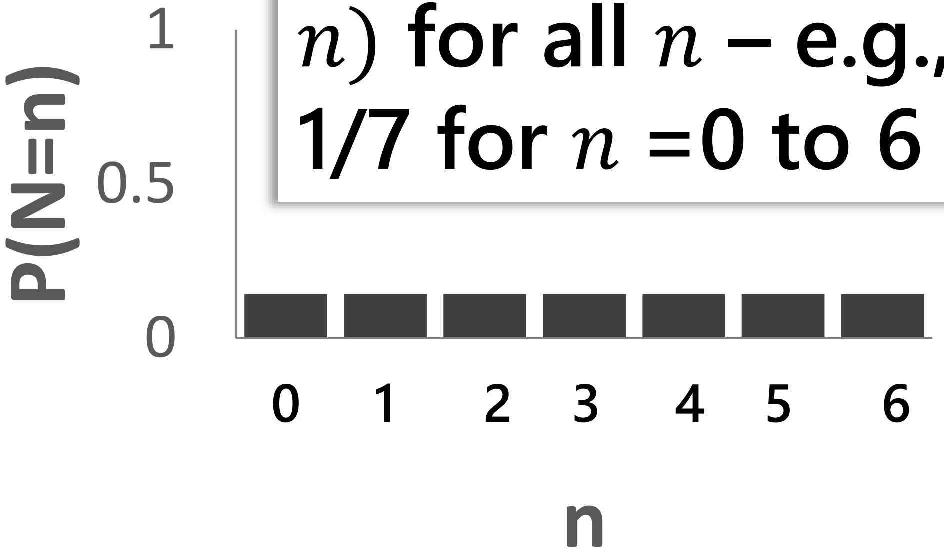
$$P(N=n|E) = \frac{P(E|N=n)\cdot P(N=n)}{\sum_n P(E|N=n)\cdot P(N=n)}$$

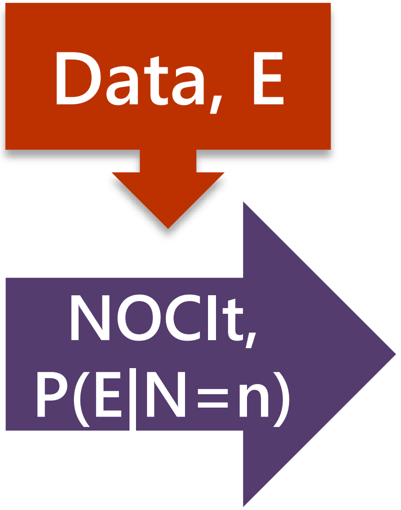
NOCIT DETERMINES $P(N=n|H_D,I,E)$

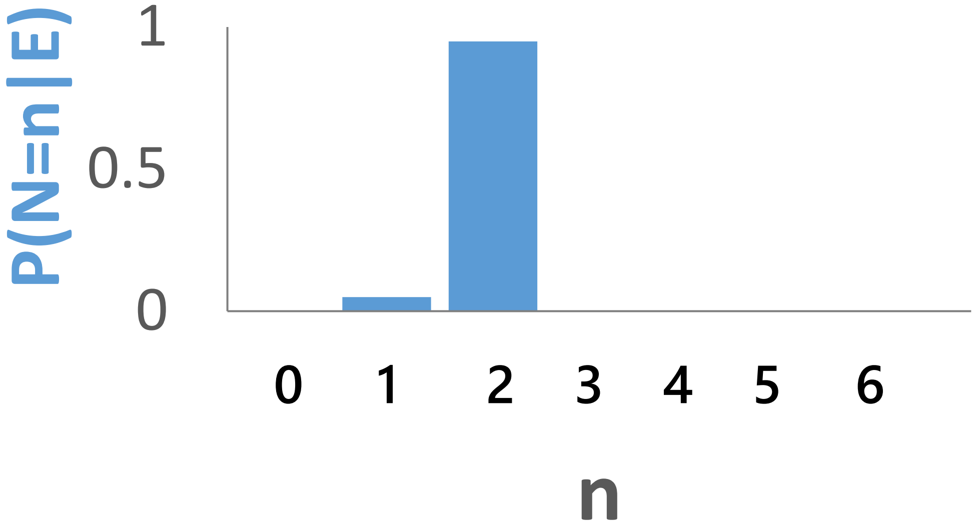
What is $P(N=n|E)$ graphically?

$$P(N=n|E) = \frac{P(E|N=n)\cdot P(N=n)}{\sum_n P(E|N=n)\cdot P(N=n)}$$

Before seeing data, set $P(N=n)$ for all n – e.g., 1/7 for n =0 to 6







What is P(*N=n|E*) *numerically*?

$$P(N=n|E) = \frac{P(E|N=n)\cdot P(N=n)}{\sum_n P(E|N=n)\cdot P(N=n)}$$

Illustrative example:

n	$P(N=n)$	$P(E\|N=n)$	$P(E\|N=n)\cdot P(N=n)$	$P(N=n\|E)$
0	1/7=0.143	0.00001	0.00000143	=0.0000014/0.0159 =0.00009
1	0.143	0.01	0.00143	0.09
2	0.143	0.1	0.0143	0.9
3	0.143	0.001	0.000143	0.009
4	0.143	0.000001	0.000000143	0.000009
5	0.143	0.0000001	1.43E-08	0.0000009
6	0.143	0.00000001	1.43E-09	0.000000009
			Sum=0.0159	**Sum=1**

Large-Scale Validation with Public Data

- Confirming NOCIt meets **predetermined** expectations
  - Unimodal distributions
    - The distribution should be peaked in one location
  - Precision
    - The apex of the distribution is the same for > 95% of the samples across 3 runs
  - Accuracy
    - $P(N=TrueNOC|H_d,I,E)$ >1% for at least 90% samples
      - TrueNOC was confirmed by running all dilutions as single-source and confirming signal
      - Used ReSOLVIt to set up a lab pipeline with LOD=1
  - Comparison
    - NOCIt outperformed current procedures
  - Robustness:
    - Resilient to sloping – i.e., degradation/inhibition effects
    - Works under different contexts

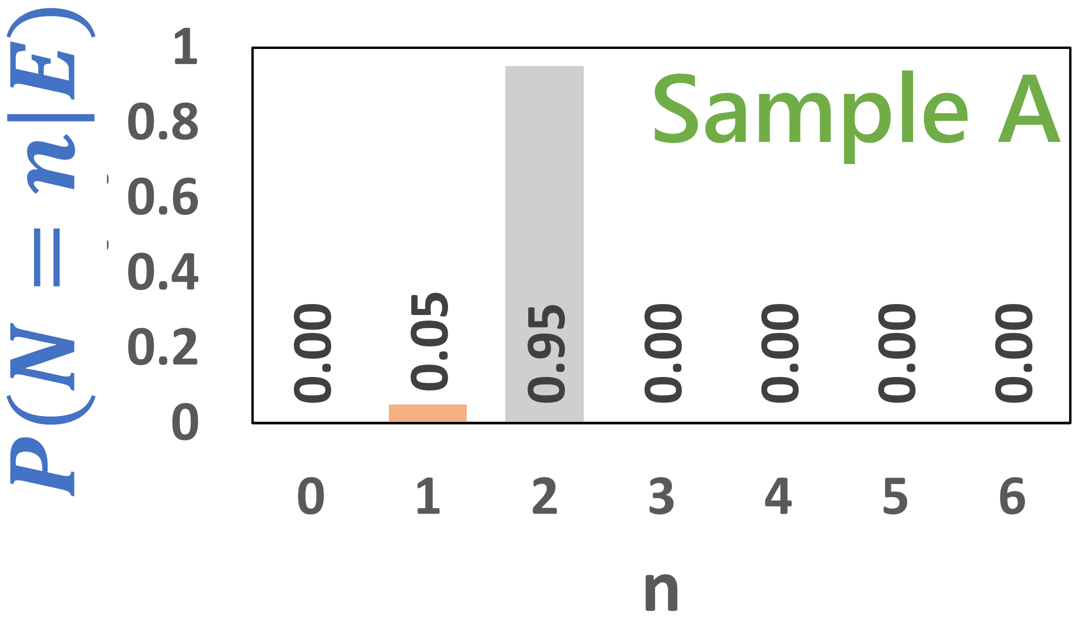
Public Mixture Data with LOD=1 are Available

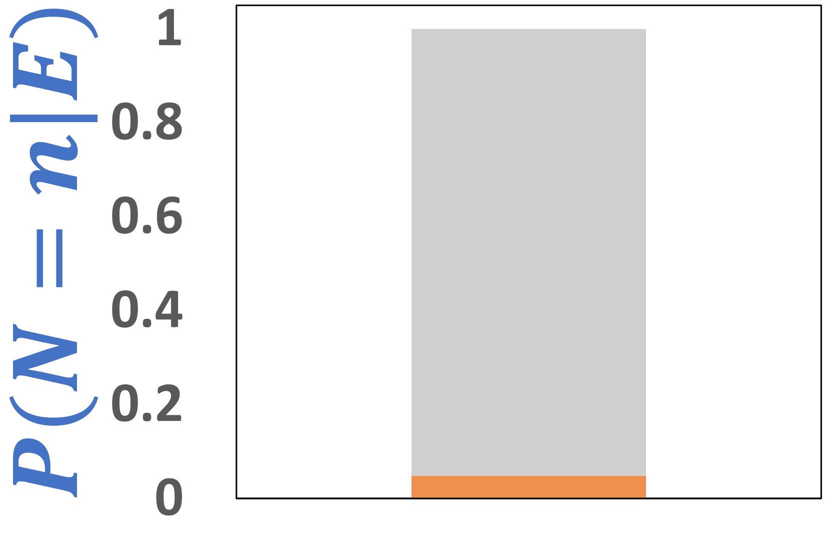
- 815 PROVEDIt samples (www.lftdi.com)
  - GlobalFiler samples (29 cycles; 25 sec 3500 Genetic Analyzer)

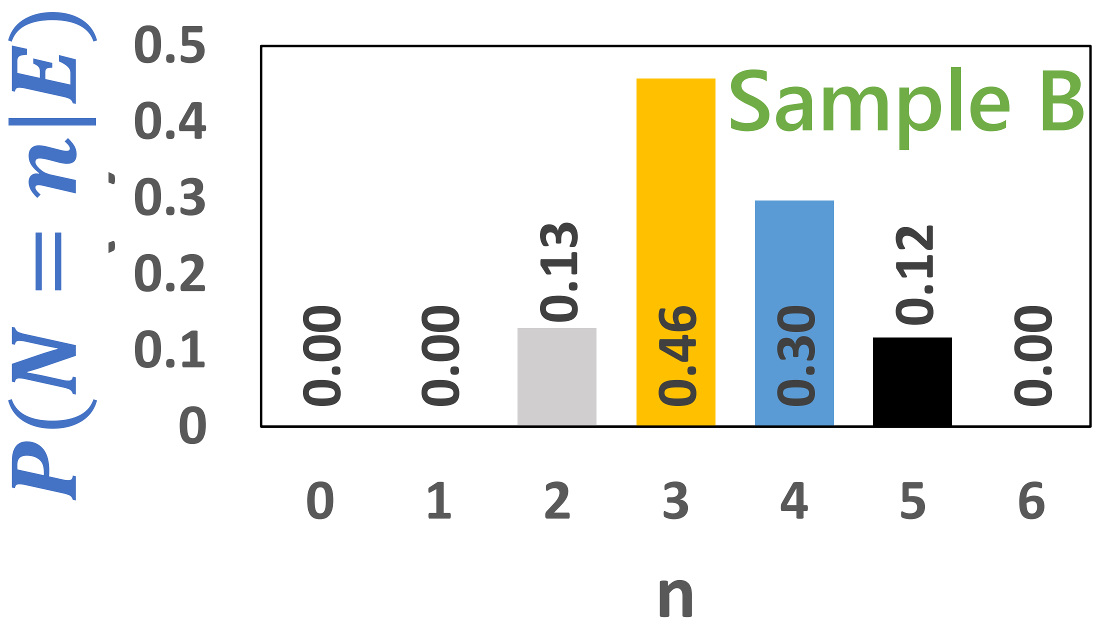
815 GlobalFiler® samples used to validate NOCIt

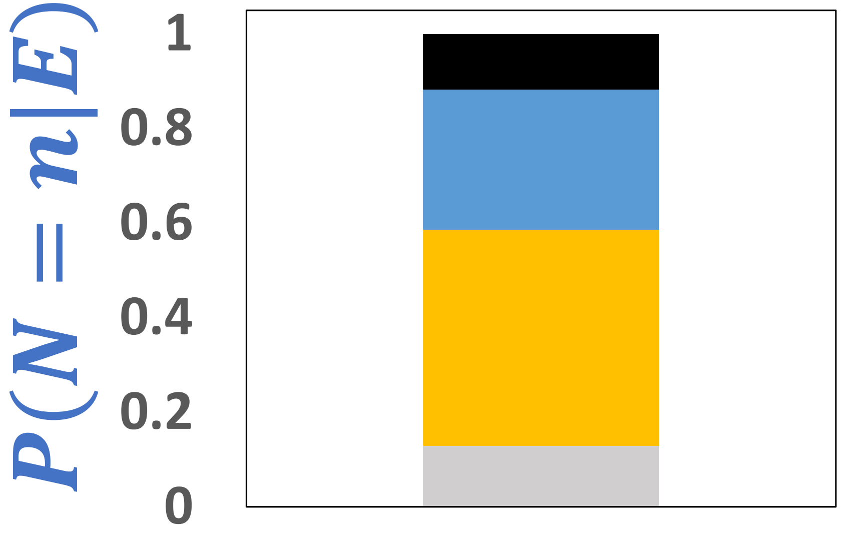
NOC	1	2	3	4	5
No. Samples	100	193	170	186	166
Tot. Template Mass (ng)	0.5 – 0.0078	0.75 – 0.03	0.75 – 0.045	0.75 – 0.06	0.75 – 0.075
Contributor Ratio	N/A	1:1 – 1:9	1:1:1 – 1:9:9	1:1:1:1 – 1:9:9:1	1:1:1:1:1 – 1:9:9:9:1

Visual Exploration is Aided by Stacked Plots

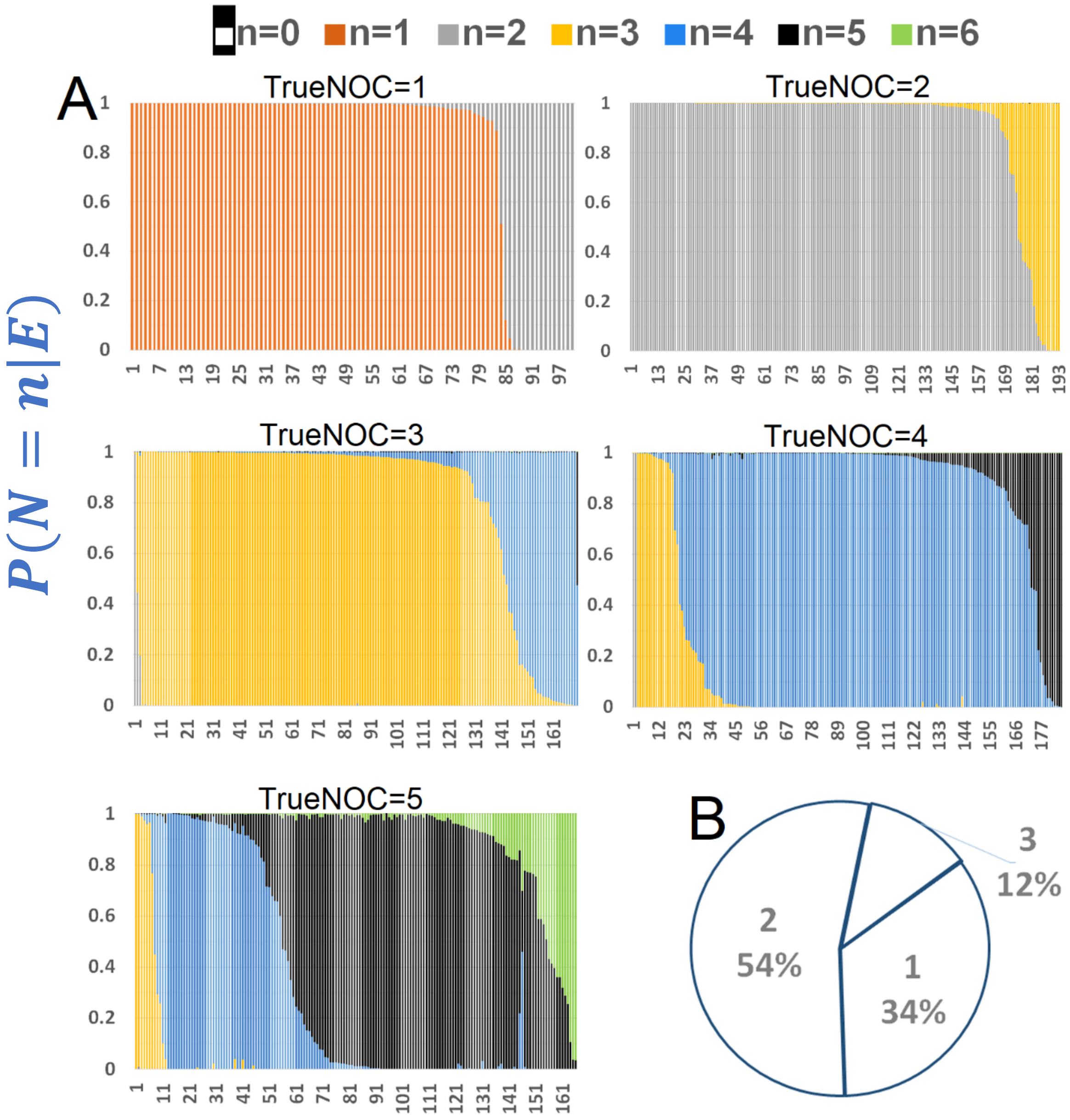








NOCIT GIVES UNIMODAL DISTRIBUTIONS



(A) Stacked Plots of APP(n) using *Condition* 1 and the APP for n= (white bar)0; ( )1; ( )2; ( )3; ( )4; and ( ) 5; and ( ) 6. X-axis is sample number.

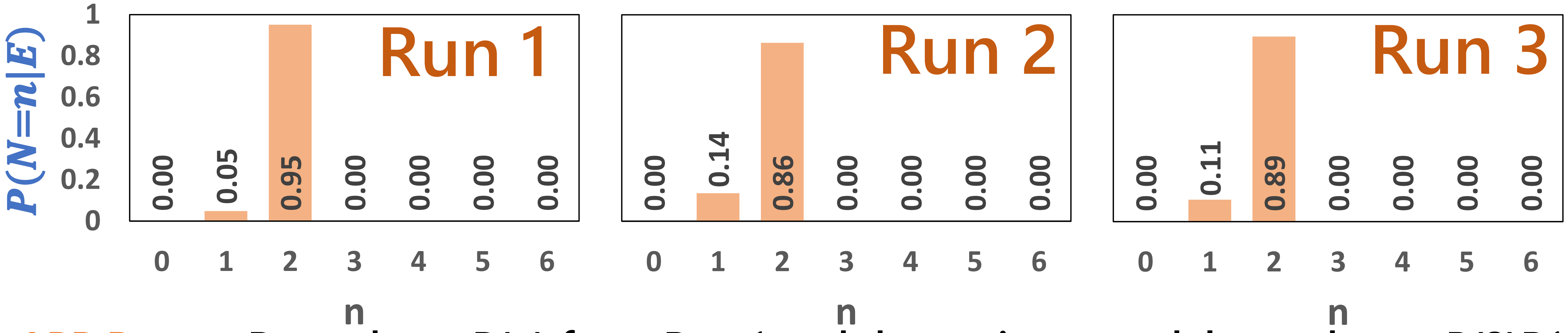
(B) Pie Chart depicting percentage of samples resulting in one, two or three APP $(n) \geq 0.001$.

Criterion: The distribution should be peaked in one location (at one n)

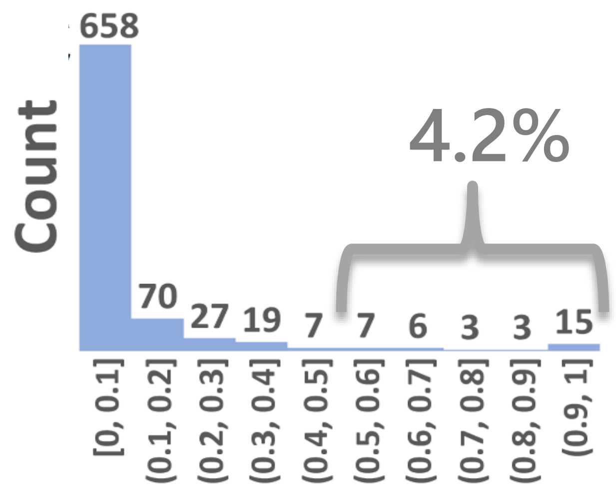
Results: Distribution was always unimodal in that there was no instance where the probability was high for small values of n, low for medium n values and then high for large n.

No sample gave more than 3 $P(N = n|E) \geq 0.001$

NOCIt is Precise Across Runs



APP Range. Record maxP(n) from Run 1 and determine max delta at that n. P(2)R1= 0.95, R2=0.86, R3=0.89. Range=(0.95-0.86)=0.09. Tells us if maxP(n) is changing.



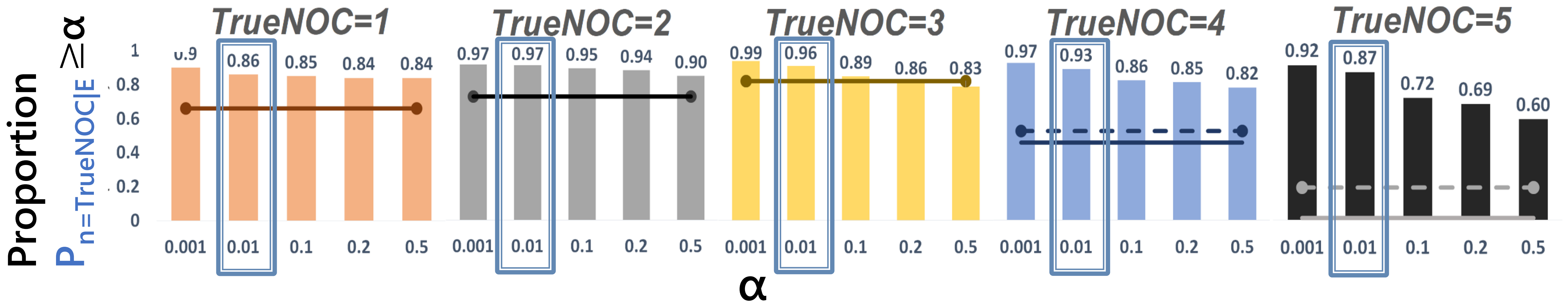
Range= max($\Delta P_{R1,R2}$; $\Delta P_{R1,R3}$; $\Delta P_{R2,R3}$)

Criterion: The distribution's apex located at the same n for > 95% of the samples across 3 runs

Results: 95.8% of the samples had repeatable distributions across 3 runs

NOCIt Is Accurate, outperforming Current Procedures

Proportion of samples giving $P(N = TrueNOC|E) \geq \alpha$, and proportion of samples for which (—) MAC and (- -) MLE (H. Haned, et al., J Forensic Sci 56(1) (2011)) estimates equaled TrueNOC



Criterion: $P(N = TrueNOC|E) \geq$ 1% for at least 90% samples

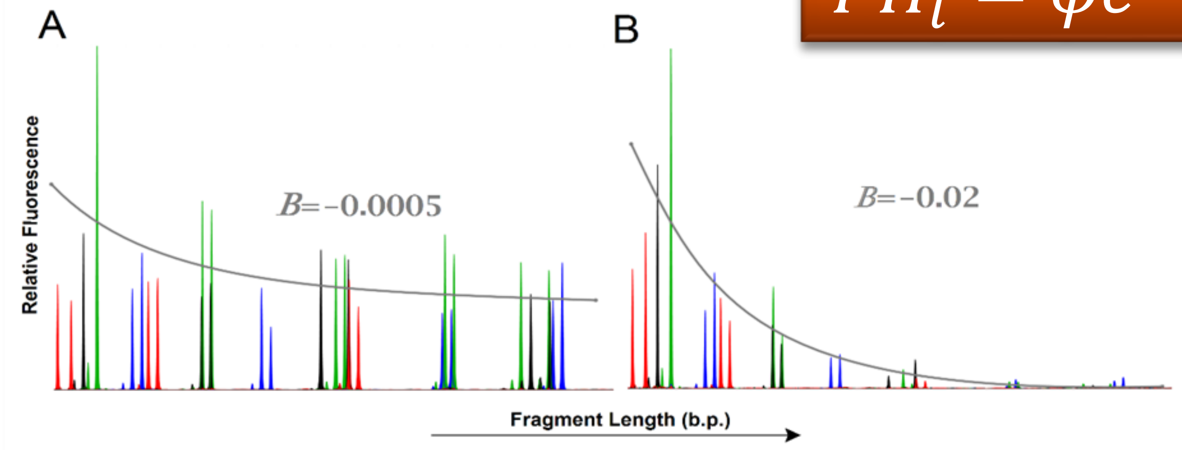
Results: 92.5% of the samples gave $P(N = TrueNOC|H_d, I, E) \geq 1\%$

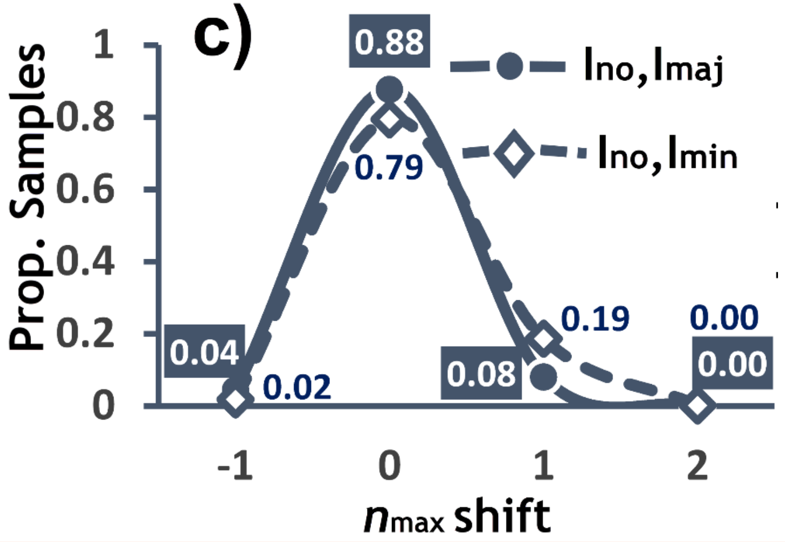
Criterion: NOCIt must outperform current procedures

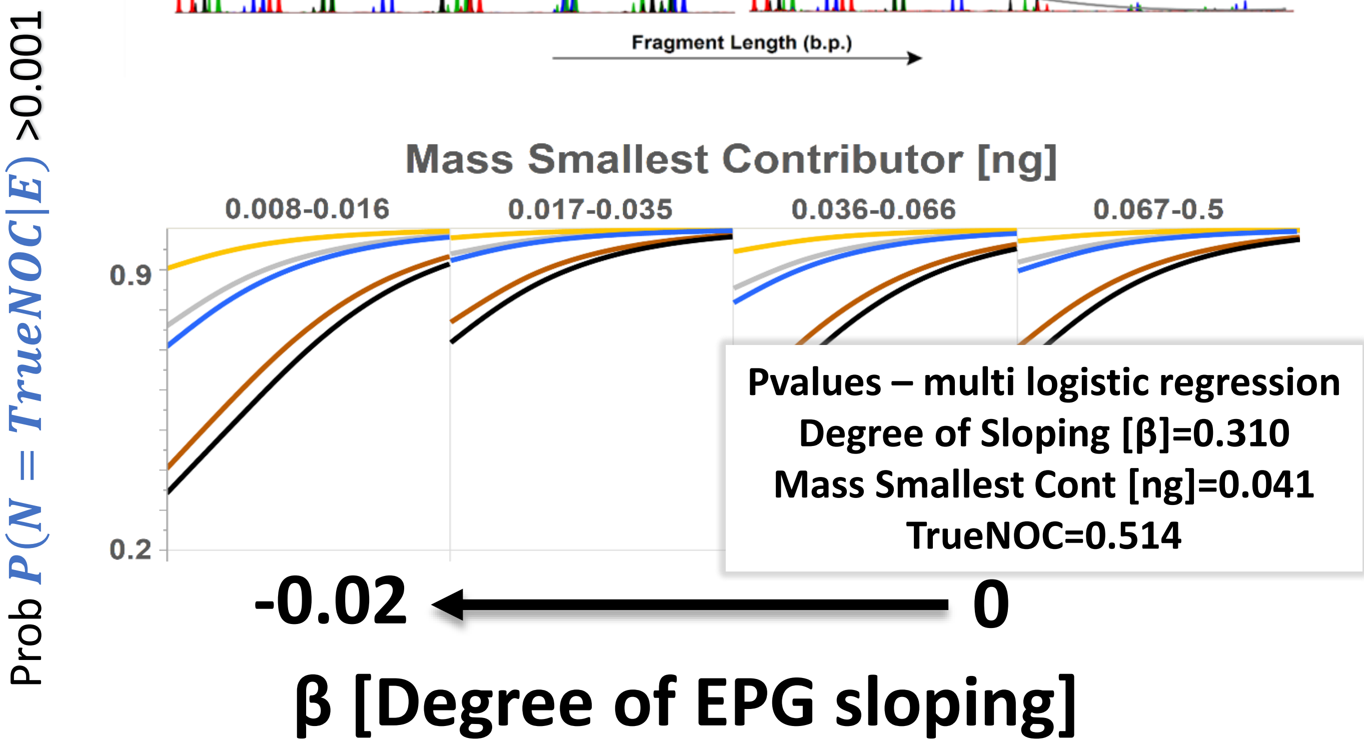
Results: NOCIt performance equals or exceeds current procedures at all α

NOCIt is Robust across Contexts and DNA Qualities

$$PH_l = \varphi e^{\beta \overline{w_l}}$$







Pvalues – multi logistic regression
Degree of Sloping [β]=0.310
Mass Smallest Cont [ng]=0.041
TrueNOC=0.514

***Criterion*: Resilient to sloping – i.e., degradation/inhibition effects**

***Results*: Pvalue of 0.310 suggests sloping [β] does not significantly affect probability of including TrueNOC when α=0.001 (or 0.5)**

***Criterion:* Works under different contexts**

***Results*: Most apexes shifted when a known contributor was assumed and it was the minor**

SUMMARY

NOCIt reports $P(N = n|E)$ – i.e., the posterior distribution for all n up to 6

Supports pipelines that do not apply default or subjective n, or automatic $P(N = n|E)$

Helps target n with non-negligible $P(N = n|E)$

Engineered to use all data (even noise)

Full descriptions in:

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

ELSEVIER

Research paper

A large-scale validation of NOCIt's *a posteriori* probability of the number of contributors and its integration into forensic interpretation pipelines

Catherine M. Grgicak[a,b,*], Slim Karkar[b], Xia Yearwood-Garcia[c], Lauren E. Alfonse[c], Ken R. Duffy[d], Desmond S. Lun[b,e,f]

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

ELSEVIER

Research paper

The a posteriori probability of the number of contributors when conditioned on an assumed contributor

FUNDING, COLLABORATORS & STUDENTS

This work was partially supported by **NIJ2011-DN-BX-K558**; **NIJ2012-DN-BX-K050**; **NIJ2014-DN-BX-K026** awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice and **W911-NF-14-C-0096** from the Department of Defense. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not reflect those of the Departments of Justice or Defense.

c.grgicak@rutgers.edu

lftdi.com

softgenetics.com

Desmond S. Lun
Slim Karkar
Ken R. Duffy
Lauren E. Alfonse
Amanda Garrett
Muriel Medard
Xia Yearwood-Garcia
Harish Swaminathan